MECHANISMS CONTROLLING HUMAN NOROVIRUS POLYPROTEIN PROCESSING

A Dissertation
submitted to the Faculty of the
Graduate School of Arts and Sciences
of Georgetown University
in partial fulfillment of the requirements for the
degree of
Doctor of Philosophy
in Microbiology and Immunology

By

Jared P. May, B.S.

Washington, DC
April 20, 2015
MECHANISMS CONTROLLING HUMAN NOROVIRUS POLYPROTEIN PROCESSING

Jared P. May, B.S.

Thesis Advisor: Dr. Brent E. Korba, Ph.D.

ABSTRACT

Noroviruses (NoV) are the prevailing cause of nonbacterial acute gastroenteritis worldwide and pose a significant financial burden on healthcare systems. The human NoV ORF1 encodes a 200 kDa polyprotein which is cleaved by the viral 20 kDa 3C-like protease (Pro, NS6) into 6 non-structural proteins necessary for viral replication. NoV ORF1 polyprotein is processed in a specific order with ‘early’ sites (NS1/2-3, NS3-4) being cleaved rapidly, and three ‘late’ sites (NS4-5, NS5-6, NS6-7) processed subsequently and less efficiently. The mechanisms controlling polyprotein processing order have not been established and we sought to determine the factors controlling this process. We have bacterially expressed and purified GI.1 Pro, GII.4 Pro, and GII.4 ProPol and characterized their buffer, pH, ion, and detergent requirements. Using short, FRET peptides representing the 5 polyprotein cleavage sites of ORF1, we have determined the enzyme kinetics of Pro and ProPol which directly correlated with the observed processing order in vitro. Enzyme turnover, $k_{cat}$, was the primary determinant of enzyme efficiency while binding affinity $K_m$ showed modest influence. We have demonstrated that ProPol is equivalent or superior to mature Pro in cleaving all ORF1 cleavage site peptides, implicating ProPol as a capable protease in the viral lifecycle. Previously, the core sequence of amino acids surrounding the scissile bonds responsible for governing the relative processing order had not been determined. Using both FRET-based peptides and full-length NoV polyprotein, we have successfully demonstrated the core sequences spanning positions P4-P2’ surrounding the NS2-3,
NS4-5, NS5-6, and NS6-7 cleavage sites contain all of the structural information necessary to control processing order in *in vitro* translation assays. We also provide insight into a previously overlooked role for the NS2-3 P3 residue in enzyme efficiency. Our models predict that the favorable electrostatic and hydrogen-bonding interactions formed between the side chains of the P3-Glu, Lys-162 of the protease, and P1-Gln residues promote the formation of a productive enzyme:transition-state complex, thus increasing enzyme turnover. Our work provides significant additional insight into understanding viral polyprotein processing and has important implications for improving the design of inhibitors targeting the NoV protease.
ACKNOWLEDGEMENTS

I would like to thank my parents, Paul and Kimberly May, for their constant support throughout my studies. I would like to especially thank my fiancée Sarah Myers for her loving support as well as listening to my research talks over the past few years.

I would also like to thank Dr. Prasanth Viswanathan for all of his technical help and guidance throughout my time at Georgetown University. My projects would not have been possible without his efforts. Next, I would like to thank Alexei Medvedev for his support and help in keeping the laboratory functional.

Next, I would like to thank my committee members Dr. Kenneth Ng, Dr. William Fonzi, Dr. John Casey, and Dr. R. Padmananabhan as well as department chairman Dr. Richard Calderone for their advice and direction throughout my time at Georgetown.

Finally, I would like to thank my mentor, Dr. Brent Korba, for supporting my work, and most importantly, allowing me the freedom to establish my own ideas and projects which has been invaluable in helping me develop as a scientist.

MANY THANKS,

JARED MAY
TABLE OF CONTENTS

Introduction ................................................................................................................................. 1
  Classification and genome characteristics ................................................................. 2
  Lack of tissue culture system ....................................................................................... 3
  Non-structural proteins ................................................................................................. 5
  NoV protease catalysis and specificity ........................................................................ 8
  Polyprotein processing ................................................................................................. 11
  Viral precursor proteins ............................................................................................... 12
  In vitro protease assay ................................................................................................. 15
  Protease inhibitors as antivirals .................................................................................... 16
  Aim of research ............................................................................................................... 16

Chapter I - Isolation and characterization of human norovirus proteases using an in vitro FRET-based assay ........................................................................................................ 18
  Results ........................................................................................................................... 19
    Cations and buffer composition reduce NoV Pro activity ........................................ 19
    Cations affected the rate of reaction but not affinity for the peptide substrate .......... 23
    Chymostatin is a potent inhibitor of human NoV protease ..................................... 25
  Discussion ...................................................................................................................... 26
  Materials and Methods ................................................................................................. 27

Chapter II - Proteolysis of peptides representing ORF1 cleavage sites by Pro and ProPol ......................................................................................................................... 31
  Results ........................................................................................................................... 32
    NoV Pro and ProPol share similar pH profile and cation sensitivities ............... 32
    Processing of ORF1 cleavage site peptides by NoV Pro and ProPol ............... 36
    Kinetic Analysis of ORF1 peptides ........................................................................... 40
    Competition of ‘early’ and ‘late’ peptides ............................................................... 42
  Discussion ...................................................................................................................... 43
  Materials and Methods ................................................................................................. 47

Chapter III - Defining the core amino acid sequence controlling polyprotein processing order ......................................................................................................................... 51
  Results ........................................................................................................................... 52
    NoV protease enzyme efficiency is controlled by the P4-P2’ residues surrounding individual cleavage sites ................................................................. 52
    Self-processing order of NoV ORF1 polyprotein can be changed by altering P4-P2’ residues ...................................................................................... 58
    NS2-3 P3 residue is important for NoV Pro enzyme efficiency ............................ 63
  Discussion ...................................................................................................................... 64
  Materials and Methods ................................................................................................. 70
Chapter IV - Effects of altered human norovirus polyprotein processing on genome replication using a transient luciferase reporter replicon ............................................................. 75
Results ............................................................................................................................. 78
  Validation and optimization of NoV reporter replicon system ......................... 78
  Effects of decreased processing at ‘early’ sites on genome replication .......... 82
  Effects of enhanced processing at ‘late’ sites on genome replication .... 85
  The second amino acid of the polymerase is important for genome replication ................................................................................................................................. 86
  Effects of preventing cleavage at individual ORF1 cleavage sites .... 87
Discussion ........................................................................................................................ 88
Materials and Methods ................................................................................................. 91
Conclusion .......................................................................................................................... 94
Appendix I– Supplementary data ............................................................................. 95
Appendix II– Copyright clearances .......................................................................... 100
Cited Literature .............................................................................................................. 106
INTRODUCTION

Noroviruses (NoV) are important human pathogens and pose a major threat to public health. NoV are single stranded, positive-sense RNA viruses that contain a naked capsid and cause acute gastroenteritis in humans (1-4). NoV are typically transmitted by the fecal-oral route through contaminated water and food but can also be aerosolized as a result of vomiting (5). NoV have been referred to as ‘cruise ship viruses’ from well-publicized cruise ship outbreaks although these outbreaks account for less than 1% of all outbreaks worldwide and only average 27 outbreaks annually (6, 7). The symptoms and illness associated with NoV infections is often referred to as ‘stomach flu’, despite little resemblance to the influenza virus. In the United States, NoV cause 21 million infections which lead to over 70,000 hospitalizations annually and account for 10-15% of severe cases of gastroenteritis in children less than 5 years old (8). NoV have also been determined to be the most frequent cause of hospital-acquired infections, leading to 65% of hospital unit closures (9). NoVs have been shown to remain infectious in groundwater for at least 61 days (10), a significant issue, especially in developing countries where NoV are estimated to cause 218,000 deaths among children each year and 1.1 million hospitalizations (11). An emerging issue surrounding NoV is the development of chronic infections in immunocompromised individuals (12, 13). For example, patients undergoing kidney transplants were shown to shed virus for extended periods of time, up to 581 days (14), which is remarkable when compared to healthy individuals who have been shown to shed virus for 13-56 days after infection (15).
Classification and genome characteristics

The family *Caliciviridae* consists of four genera of viruses. The genera include *Norovirus, Lagovirus, Sapovirus*, and *Vesivirus*. Caliciviruses infect a variety of animal hosts, including bovine, ovine, murine, canine, feline, marine mammals, and swine, among others (16-22). The symptoms associated with calicivirus infections can vary drastically. Aside from the *Norovirus* genus, symptoms of caliciviruses infections range from upper respiratory tract infection to lesions, stomatitis, and hemorrhagic symptoms (16). *Norovirus* infections display acute gastroenteritis in humans but are mainly asymptomatic in animal hosts (4).

Within *Norovirus*, five genogroups (GI-GV) have been recognized, with the very recent addition of a sixth genogroup that infects canines (GVI) (4, 23, 24). The genogroups and subgroups are distinguished by their respective VP1 capsid sequences. The human NoV belong to GI, GII, and GIV, with the majority of outbreaks worldwide caused by GII strains, especially the GII.4 genotype (4, 25, 26). To improve reporting measures, the CDC and State partners have developed the CaliciNet reporting system and they determined that 73% of NoV outbreaks from Oct. 2009 through May 2010 were from the GII.4 genotype (27). NoV outbreaks can occur at any point throughout the year but have been reported to be concentrated in the winter months (28, 29). Due to the dominance of the GII.4 genotype, it appears that epidemics are occurring as a result of the rapid evolution of the GII.4 genotype, possibly driven by population immunity (25, 30).

The prototype human Calicivirus, Norwalk Virus (GI), is non-enveloped and has a 7.7kb positive sense single-stranded RNA genome that encodes three open reading frames (ORF1-3) (see Figure i.1) (1-4). ORF1 encodes a 200 kDa nonstructural polyprotein that is subsequently
auto-catalytically cleaved into its 6 mature proteins (2-4, 31). Like all +ssRNA viruses, membrane rearrangements are observed and NoV infection results in membrane clustering from the endosomal, trans-golgi, and endoplasmic reticulum (ER) membranes and serves as the center of the replication complex where the non-structural proteins tend to aggregate (32-34). ORF2 and ORF3 encode for the major (VP1) and minor (VP2) structural proteins respectively. VP1 is the major component of the viral capsid and has been shown to enhance viral polymerase activity, suggesting a role in the viral replication complex (35), while the role of VP2 remains unclear.

**Figure i.1: Genomic organization of human NoV.** The human NoV genome contains three open reading frames (ORF1-3). ORF1 contains the nonstructural proteins and is translated as a single polyprotein from the genomic RNA. ORF2-3 are translated from a subgenomic RNA and code for the structural proteins. VPg is covalently linked to the 5’ end of both genomic and subgenomic RNA’s and is required for efficient translation.

**Lack of tissue culture system**

The biggest hindrance to the advancement of research into NoV biology has been the lack of a tissue culture system. A single report concluded 3D tissue culture or organoid model of human small intestinal epithelium could support human NoV replication but to date has not been replicated (36). Numerous studies have attempted to find cell types that could support human NoV in tissue culture with little success (37-40). Recently, a report demonstrated replication of human NoV in B cells with enteric bacteria as a stimulatory factor (41) and has been supported
by an additional study which demonstrated commensal bacteria were required for persistent infection of murine norovirus (42). Although a robust infection system is absent, a single cycle of human NoV replication has been observed in the human HEK-293T and Huh-7 cell lines. Cotransfection of the full-length genome and subgenome under control of a T7 promoter was required to produce particles in HEK-293T cells infected with vaccinia-T7 (43). However, viral RNA extracted from NoV-challenged stool samples could produce particles after direct transfection into Huh-7 cells, eliminating the need for a helper-virus (44). An advancement in adapting human NoV for tissue culture includes a stably expressed NoV reporter replicon which has a portion of VP1 replaced with a neomycin resistance gene (45). While this replicon provides a valuable tool for testing the effects of NoV inhibitors in a cell-based assay, it cannot be manipulated as a reverse genetics system to test the effects of individual mutations on viral replication. Instead, a newly developed reverse genetics system for GII human NoV has been developed (46). This system relies on an EF-1α mammalian promoter to drive the expression of the NoV genome with luciferase inserted into ORF2 and measuring bioluminescence as a result of subgenomic RNA expression (46). Despite this significant advancement, the lack of a susceptible cell type or cell-line for human NoV precludes any studies on the infectivity of viruses generated from the GII reverse genetics system.

Caliciviruses that can be propagated in tissue culture have traditionally been used as models, including Feline calicivirus (FCV) and porcine enteric calicivirus (PEC) (47-49). More recently, Murine Norovirus (MNV) has been discovered and cultivated in tissue culture (17, 50, 51). MNV was first discovered as an interferon-sensitive pathogen in STAT1-deficient mice and caused lethal encephalitis, vasculitis, and meningitis (17). MNV was found to replicate in tissue
culture in the RAW 264.7 murine macrophage-like cells providing a robust infection system (50).

Non-structural proteins

The non-structural proteins of ORF1 are required for genome replication and are translated as a single polyprotein (see Figure i.2). The ORF1 polyprotein has been demonstrated to induce apoptosis through caspase-9 activation, highlighting another important role for this polyprotein (52). Standard nomenclature has been suggested to designate the proteins as (NS1-NS7) starting from the N-terminus of the polyprotein (53, 54). At the N-terminus, p48, or Nterm (NS1/2) is one of the least studied NoV proteins and lacks any defined structure which has made predictions of its function difficult (55). NS1/2 has been shown to be involved in disrupting intracellular protein trafficking (56). Also, NS1/2 has been implicated in disrupting golgi trafficking, which was dependent on a predicted hydrophobic region (57). In MNV, a single amino acid mutation in NS1/2 has been demonstrated to confer viral persistence during an infection (58).

**Figure i.2: Human NoV open reading frame 1 (ORF1).** The ORF1 polyprotein is pictured with accepted nomenclatures and predicted molecular weights for mature proteins. At each cleavage site (marked with asterisk) the dipeptide at the scissile bond is shown. ‘Early’ cleavage sites contain a QG dipeptide at the scissile bond while ‘Late’ cleavage sites contain either EG or EA dipeptides.
To the C-terminus of p48 is p41, or NTPase (NS3). NTPase has been demonstrated to possess nucleoside triphosphatase activity (59). Although NTPase shares sequence homology to the 2C helicase of picornaviruses, no helicase activity has been detected to date (59).

The p20 (NS4) protein lies to the C-terminus of p41 and is the most poorly understood of the non-structural proteins. NS4 contains a membrane association domain and has been speculated to serve as a membrane anchor for the replication complex (60) and is supported by the finding that NS4 interacts with NS1/2 which has been shown to induce membrane rearrangements (61). NS4 has also been implicated in disrupting the actin redistribution pathways by inducing expression of NS4 in the HT-29/B6 epithelial cell line (62).

To the C-terminus of p20 lies p15, referred to as VPg (NS5). VPg (Viral Protein, genome-linked) is a small protein that is covalently linked to the 5’ end of the viral genome (4). The covalent attachment of VPg to the viral genomes is necessary for replication as uncapped RNA is not infectious in calicivirus reverse genetics systems. Substitution with a m7G(S’)ppp(S’)G cap analog in non-human caliciviruses permits initiation of infection and subsequent replication (63-66). Human NoV RNA is not infectious with a 5’ cap analog which has been a major roadblock to the development of a human NoV reverse genetics system. The 5’ cap analog could be sufficient for RNA infectivity in other caliciviruses since it could function as a pseudo-VPg in the recruitment of the eukaryotic translation machinery (eIF3, eIF4GI, eIF4E) as several studies have implicated VPg in this process (67-70). The exact mechanism of the covalent attachment of VPg to the 5’ end of the viral genome is unknown. However, VPg is thought to be involved in genome replication through nucleotidylation by the viral polymerase and then serve as a protein primer for viral transcription (71-73). The nucleotidylation of VPg is
critical for replication as mutagenesis of the target residue for nucleotidylation abolished viral replication (74).

The viral protease, Pro (NS6) lies next to the C-terminus of VPg (Figure 1). Pro functions to cleave the ORF1 polyprotein into its mature proteins that have been demonstrated to be essential for replication (75). Sequence alignments were used to identify the viral protease and the homology to the 3C proteases of picornaviruses (76, 77). Although it has yet to be demonstrated for NoV proteases, a role in antagonizing the innate immune response could exist for Pro. In hepatitis C virus (HCV), the NS3 protease is responsible for subverting the innate immune response by cleaving the mitochondrial antiviral signaling protein (MAVS) and blocking interleukin 28 (IL-28) synthesis (78, 79). In Enterovirus 71, a picornavirus, the 3C protease has been shown to inhibit retinoid acid-inducible gene I (RIG-I), which prevents RIG-I from triggering an interferon (IFN) immune response (80). Another example of a 3C protease antagonizing the host innate immune response is Coxsackievirus B3 3C Pro cleaving the mitochondrial antiviral signaling protein (MAVS) and Toll/IL-1 receptor domain-containing adaptor inducing interferon-beta (TRIF) which are critical proteins for inducing a type-I IFN response (81). Finally, in foot-and-mouth disease virus (FMDV), the 3C Pro has been demonstrated to cleave NEMO, a protein required for NF-kβ function and downstream IFN pathways (82).

The most C-terminal protein in ORF1 is the viral RNA-dependent-RNA-polymerase (RdRp). The NoV RdRp, or Pol, has had its structure solved, and is a 56 kDa “right hand” polymerase responsible for replicating the viral genome by using a negative-sense intermediate RNA molecule (83, 84). The NoV RdRp has been biochemically characterized and has been
shown to possess an error rate of $1 \times 10^{-4} – 1 \times 10^{-5}$ in vitro (85, 86). Pol has been shown to nucleotidylate VPg in vitro as well as synthesize RNA’s de novo and in the presence of an oligonucleotide primer (72, 86, 87). The polymerase has been shown to generate antigenomic RNA in a primer-dependent manner in vitro, and VPg-polyU is thought to serve as the primer (72). Since RNA replication is critical for viral fitness and other viral polymerases (e.g. HCV) have been successfully targeted (88), NoV Pol has been targeted for antiviral development (89-91).

*NoV protease catalysis and specificity*

The structures of 3C-like proteases of the Chiba, Southampton, Murine, and Norwalk noroviruses have been solved and the amino acid residues important for substrate interactions have been designated (92-96). The overall structure of NoV Pro is that of a cysteine protease with a chymotrypsin-like fold and a catalytic triad consisting of His-30, Glu-54, and Cys-139 (93, 97). The structures of NoV proteases share a high degree of structural similarity to picornaviral 3C proteases (98, 99), making picornaviral proteases an excellent model for comparison. The mechanism of catalysis for the NoV proteases is similar to the serine proteases that contain a Ser-His-Asp triad in the active site. The mechanism of catalysis of serine proteases is shown in Figure i.3. The catalysis reaction occurs in two halves, acylation and deacylation. First, the carbonyl of the peptide substrate is attacked by the Serine and His residues of the catalytic triad (100). The result is the first of two tetrahedral intermediates which collapses to form the acylenzyme intermediate (100). The second half of the reaction is very similar where a water molecule and His residue attack the acylenzyme intermediate to yield the second “transition-state” intermediate which collapses to release the carboxylic acid product (100).
Similar to the serine proteases, the Cys-139 residue of NoV proteases functions similarly to the Ser residue and the Glu-54 functions as the Asp residue of the serine proteases while the His residue function is conserved.

**Figure i.3: Mechanism of catalysis of serine proteases.** Reprinted (adapted) with permission from (Hedstrom L. 2002. Serine protease mechanism and specificity. Chemical reviews (102:4501-4524.). Copyright (2002) American Chemical Society.

Utilizing crystallographic structures of the NoV proteases, insights into the substrate specificity of the protease have been described. Multiple structures of Pro with substrate peptides has indicated the preferred conformation of substrate is an extended β-strand which can use hydrogen bond networks to form antiparallel β-sheets with the β-strands of the protease (94-96). These networks of hydrogen bonds between the substrate and the protease are thought to be the driving force for correctly positioning the substrate for proteolysis. Using nomenclature described by Schechter and Berger (1967), the substrate amino acids are named P1, P2, P3, etc. from the scissile bond towards the N-terminus while the amino acids from the scissile bond towards the C-terminus are named P1’, P2’, P3’, etc. (101). For NoV proteases as well as 3C proteases of picornaviruses and coronaviruses, the P1 amino acids is either Glu (E) or Gln (Q)
The P2 position is always occupied by a hydrophobic side chain such as Leu (L), Met (M), or Phe (F) (31). P3 has not been implicated in substrate specificity of the protease, largely due to the diversity of amino acids found in this position in ORF1 cleavage sites (31, 94). The P4 position of the substrate is generally hydrophobic including Phe (F), Ala (A), Ile (I), or Thr (T) (31). Amino acids at P5 and positions towards the N-terminus are not thought to play a role in substrate specificity, but the presence of an amino acid at P5 can increase the cleavage efficiency of small peptides in *in vitro* protease assays (94, 106). The available structural data has provided little insight into the role of amino acids to the C-terminus of the scissile bond in substrate specificity, largely due to the lack of structures that contain peptides with residues to the ‘prime’ side of the scissile bond. As a result, any information regarding the role of ‘prime’ side residues in substrate specificity is speculative.

Examination of the crystal structures of NoV proteases in the presence of peptides or inhibitors has indicated that the substrate residues P4, P2, P1, and P1’ have substantial interactions with the protease during proteolysis (94-96, 107, 108). The P2’ residue has been hypothesized to direct the substrate into the “surface cleft” of the viral protease using the disruptive residues P1’-Gly and P2’-Pro (94). Furthermore, the P2’ residue has been demonstrated to modulate cleavage efficiency of the related sapovirus 3C-like protease (109). To date, the P3 residues surrounding NoV ORF1 cleavage sites have not been implicated in substrate specificity or enzyme efficiency by *in vitro* processing (110) or crystallography studies (94-96), similar to observations for picornaviral 3C proteases (111, 112).
Polyprotein processing

The human NoV ORF1 encodes a ~200 kDa polyprotein which contains the nonstructural proteins required for replication. Polyproteins are common among positive-sense RNA viruses and allow the virus to express many genes from a single RNA molecule while condensing the genetic material necessary for replication. This is necessary since host cells translate mRNAs that code for a single protein. The viral protease, Pro (NS6) is responsible for cleaving the polyprotein into the 6 mature proteins needed for replication. The protease is thought to be able cleave the polyprotein both in cis and in trans, however, true cis-cleavage is difficult to demonstrate experimentally. Earlier work has defined the amino acid dipeptide sequences at which the specific protease cleavages occur (31, 53, 113-115). The two N-terminal cleavage sites (NS1/2-3 and NS3-4) of human NoV are cleaved most efficiently in in vitro assays and have a QG dipeptide at the scissile bond that undergoes cleavage by Pro (see Figure i.2) (109, 113, 116, 117). As a result of this efficient cleavage, mature Nterm (NS1/2) and NTPase (NS3) are observed early in in vitro translation assays (113, 116-118). These sites have been termed ‘early’ cleavage sites. The remaining cleavage sites (NS4-5, NS5-6, and NS6-7) are cleaved much less efficiently by Pro and are referred to as ‘late’ cleavage sites (113, 116-118). The NS4-5 and NS6-7 cleavage sites contain an EG dipeptide at the scissile bond while the NS5-6 cleavage site contains an EA dipeptide for the human NoV (see Figure i.2). Earlier work on ORF1 polyprotein processing had proposed a specific processing order of the polyprotein by Pro but the mechanisms controlling the order of cleavage events were not determined (113).
**Viral precursor proteins**

As a result of the inefficient cleavage of the ‘late’ sites by Pro, polyprotein precursor proteins are observed both in vitro and in vivo for human NoV and MNV (46, 53, 113). Nterm and NTPase-containing precursors are not observed due to the efficient cleavage at the NS1/2-3 and NS3-4 cleavage sites. However due to the inefficient cleavage at the ‘late’ sites, a 114 kDa precursor, p20VPgProPol, is observed along with other variations such as p20VPg and ProPol (53, 113).

Of interest, the 3CD precursor (picornaviral ProPol homolog) has been described to function differently than mature 3C (Pro) in picornaviruses. In picornaviruses, the activity of the mature 3C protease differs from 3CD since 3CD preferentially processes the P1 precursor (119-121). Furthermore, 3CD precursor, but not mature 3C has been demonstrated to efficiently cleave the VP1/2A polyprotein cleavage site in the Aichi virus (122). In the picornaviral lifecycle, the protease 3C has been shown to shut-off host-cell transcription by cleavage of transcription factors (123). For 3C Pro to be able to cleave transcription factors, it must be able to enter the host cell nucleus. Data suggests the 3CD precursor contains a nuclear localization signal in the 3D (Pol) domain, which is responsible for transporting 3C to the nucleus where it can undergo self-cleavage and shut off host-cell transcription (124, 125). Unlike human NoV, the 3CD precursor of picornaviruses does not exhibit polymerase activity due to inhibition, while the NoV ProPol has been shown to possess both protease and polymerase activities simultaneously (126-128). For Poliovirus, it has been proposed that the viral polymerase becomes active after cleavage of 3CD as a result of N-terminal residues rearranging following proteolysis (129). The 3CD precursor of Hepatitis A virus (HAV), also a picornavirus, has been implicated in
antagonizing immune signaling of host cells by cleaving TRIF, an adaptor protein of the TLR3 signaling pathway (130). The 3D (polymerase) domain was suggested to modulate the substrate specificity of the 3C protease domain allowing it to recognize atypical cleavage sequences in TRIF (130).

The 3CD precursor of poliovirus has been shown to influence a switch from viral genome translation to viral transcription (131). For positive-sense RNA viruses, the genome can be directly translated once released in the cytoplasm, however antigenome transcription cannot occur simultaneously due to the translation machinery hindering the accessibility of proteins required for transcription. A host protein at the center of the process of switching from translation to transcription is the poly(rC) binding protein 2 (PCBP2). PCBP2 up-regulates viral genome translation when intact, but upon cleavage PCBP2 is associated with antigenome synthesis and multiple studies have shown that 3CD is capable of cleaving PCBP2 (132-134). Finally, 3CD has been shown to bind the viral genome near the 5’ end and has been shown to bind a cis-acting RNA replication element, or cre, which enhanced initiation of positive-strand synthesis or uridylylation of VPg respectively (135, 136).

Both human NoV Pro and ProPol possess proteolytic activity and are capable of processing the ORF1 polyprotein (109, 113, 116, 117, 137). Both Pro and ProPol have been observed to cleave a NoV reporter, which consisted of Gaussia Luciferase fused to NTPase with a NoV cleavage sequence, in a cell-based assay (138). It has been proposed that Pro and ProPol may differentially cleave the ORF1 polyprotein (109, 113, 137), but there is limited data on the extent to which ProPol processes the ‘late’ sites of the ORF1 polyprotein.
The NS4-5 precursor, p20VPg, has been detected in vesivirus and MNV infected cells as a stable precursor protein, likely as result of the inefficient cleavage of the NS4-5 ‘late’ site (53, 75). The function of the p20VPg (NS4-5) precursor in the NoV lifecycle is not known. In picornavirus, the p20VPg homolog, 3AB, is present during an infection and is thought to be a functional form of precursor VPg in uridylylation (139-141). The 3AB precursor of picornaviruses has also been described to exhibit nucleic acid chaperone activity (142, 143), highlighting yet another role of this viral precursor protein in replication. In addition to 3AB, the 3BC precursor of picornavirus (similar to VPgPro of NoV) is thought to be a substrate for uridylylation of VPg since abolition of cleavage between 3B and 3C was not lethal and resulted in 3BC-linked RNA (144, 145). It has also been hypothesized that precursor forms of VPg are the main substrate for uridylylation compared to mature VPg and is supported by in vitro data which demonstrated 3BC (VPgPro) was the most kinetically favorable donor of VPg for uridylylation (145).

The roles of precursor proteins in viral replication of NoV are not well understood. To determine the function of precursor proteins, it is possible to prevent the formation of precursors by enhancing the cleavage between proteins of proteins of interest. This has been demonstrated in HCV where amino acid substitutions which enhanced cleavage at the NS4B/5A boundary were lethal; indicating the precursor protein involved must exist for a finite amount of time (146). In hepatitis A virus (HAV), increased processing at the 3AB site decreased viral replication, further demonstrating the importance of long-lived precursor proteins in viral replication (147). In order to enhance the cleavage of NoV cleavage sites, the core sequence of
amino acids surrounding the cleavage sites that control enzyme efficiency must be fully elucidated.

_In vitro protease assay_

The NoV protease is considered to be one of the premier antiviral targets and numerous compounds have been developed to inhibit protease activity. A critical tool to evaluate the efficacy of experimental protease inhibitors is a robust and high-throughput _in vitro_ protease assay. A number of assays have been developed for 3C proteases of the picornaviruses and coronaviruses, many of which use 5(6)-Carboxyfluorescein or fluorescence resonance energy transfer (FRET) technology (148-151). FRET-based protease assays incorporate different donor and acceptor dyes on the N-terminus and C-terminus of a peptide substrate. When intact, the fluorescence of the donor is quenched by the acceptor due to their close proximity, however upon cleavage, the donor and acceptor dyes are released in separate fragments of the peptide substrate and an increase in fluorescence is observed (152, 153). FRET-based protease assays have been used for NoV proteases using the Edans/Dabcyl dye combination (93, 154, 155). Using a microplate reader with excitation and emission wavelengths of 360 nm and 460 nm respectively, a simple readout of relative fluorescence units (RFU) is measured in 96-well format, allowing for high-throughput screening. These protease assays were commonly used to determine the protease inhibitory effects of various compounds but the NoV protease assay was not optimized to achieve the highest signal-to-background ratio or characterize the protease by determining its optimum pH or buffer conditions.
**Protease inhibitors as antivirals**

While much of the field has focused on the development of a NoV vaccine (156, 157), antiviral development has been actively pursued (13). The human NoV protease has been the most targeted non-structural viral protein for antiviral drug discovery to date (13, 158). Numerous studies have identified compounds that exhibit protease inhibition in *in vitro* as well as in a cell-based assay using a NoC reporter replicon (45). A common strategy employed to target viral proteases includes using inhibitors that ‘mimic’ the natural substrate of the protease. Such inhibitors that have been found to be effective against the human NoV protease include peptidyl aldehyde inhibitors (159), α-ketoamide and α-ketoheterocycle compounds (160), a dipeptidyl bisulfite adduct salt compound (107), dipeptidyl α-hydroxyphosphonates (161), and a ketomethylene-containing peptidomimetic compound (162). The substrate specificity of the human NoV protease must be studied in great detail to provide valuable information for intelligent drug design for protease inhibitors. It is well known that NoV proteases prefer glutamine (Q) at P1 and hydrophobic residues at P2 and has been taken advantage of in the design of previous protease inhibitors. However, the substrate specificity, if any exists, at P3 has not been elucidated and could provide additional insight to improve the efficacy of peptidomimetic inhibitors.

**Aim of research**

The mechanisms controlling NoV polyprotein processing order have not been fully elucidated. Furthermore, the functions of precursor proteins in the NoV lifecycle are not well understood. Therefore, the aims of these studies were as follows:
i. Purify recombinant GII Pro and GII ProPol and characterize their biochemical properties by using an improved FRET-based protease assay

ii. Determine the enzyme kinetics of Pro and ProPol using short FRET peptides representing the five natural ORF1 cleavage sites

iii. Determine the core sequence of amino acids surrounding ORF1 cleavage sites that control protease efficiency and processing order

iv. Alter polyprotein processing order in vitro by switching ‘early’ and ‘late’ cleavage sites

v. Determine the effects of altered polyprotein processing on viral replication in cell culture using luciferase-based reporter constructs.
Chapter I

Isolation and characterization of human norovirus proteases using an in vitro FRET-based assay

Reprinted in part with permission from the following manuscript:

To study the human NoV protease \textit{in vitro}, recombinant protein must be expressed and purified. In this study, we cloned, expressed, and purified both GI and GII human NoV proteases. The recombinant proteins were histidine-tagged for easy isolation and were used to characterize the optimal assay conditions for protease activity. The resulting assay is an invaluable tool which we later utilized to probe the substrate specificity and mechanisms controlling polyprotein processing.

\textbf{Results}

\textit{Cations and buffer composition reduce NoV Pro activity.}

Nearly all previously published human NoV protease (NoV Pro) assays have utilized 50-100 mM sodium phosphate buffer, pH 8.0 or greater, NaCl at 100-150 mM, enzyme concentrations of 2 µM or greater, and relatively low substrate:enzyme ratios (less than 10) (93, 94, 109, 137, 154, 163, 164). However, there is little information on the relative effect of alternative assay conditions, or a systematic examination of reaction conditions and components for a FRET-based assay. The standard buffer composition for the NoV Pro activity assay in the current study (10 mM HEPES, pH 7.6, 0.1% CHAPS, 5 mM DTT, 30% glycerol) was based on empirical observations that examined the effect of pH, detergent, buffering agent, and commonly utilized cations.

Activity of NoV Pro from both GI and GII viruses was examined and was found to be essentially identical overall. Protease activity was maximal at pH 8.6-9.0, but strong protease activity at more physiologically relevant levels was retained (Fig. 1.1A). At pH 7.6, the GI Pro retained 73% and the GII Pro retained 63% of the activity at the optimal pH (Fig. 1.1A). Protease activity in HEPES was at least 2-fold higher than in Tris-HCl or NaPO₄ across a range of pH
values (Figs. 1.1B, C). Protease activity at buffer concentrations higher than 10 mM was reduced for all three buffering agents examined at pH 7.6 (Fig 1.1C).

**Figure 1.1. Effect of pH and buffering agent on NoV Pro activity.** NoV Pro concentration was 0.5 µM and standard assay substrate (Edans-EPDFHLQGPEDLAK-Dabcyl) was held at 100-fold excess for all conditions. Values were determined at 60 minutes of reaction (near 80% completion). A) Effect of pH on NoV Pro activity. Assays were performed in duplicate with both data points shown. B) Effect of buffering agent on NoV Pro. C) Effect of buffer concentration on NoV Pro. Data are presented as a percentage of the maximum RFU value. Assays were conducted in duplicate for each condition.
For the GI Pro, NaCl inhibited activity 2-fold at approximately 90 mM, but higher concentrations up to 300 mM had little additional effect (Fig. 1.2A). KCl, MgCl$_2$, MgSO$_4$, CaCl$_2$, and MnCl$_2$ were more inhibitory, inducing a 2-fold reductions at as little as 5-10 mM in some instances, and up to 10-fold at higher concentrations (Figs. 1.2A, B). Relative to the GI protease, the GII enzyme appeared to be slightly more sensitive to NaCl and MgCl$_2$ and slightly less sensitive to MgSO$_4$ and CaCl$_2$ (Figs. 1.2A, B). Data for zinc was not included as addition at any concentration induced precipitation.

While the addition of CHAPS and DTT only modestly enhanced GI Pro activity (with a greater apparent effect on GII Pro at lower concentrations) (Figs. 1.2C, D), DTT and CHAPS were included in the final reaction buffer to potentially aid solubility and stability of the enzyme. EDTA at concentrations up to 0.5 µM, DMSO at concentrations of up to 5%, and glycerol concentrations up to 30% did not affect protease activity (data not shown).
Figure 1.2. Effect of cations and assay components on NoV Pro activity. NoV Pro concentration was 0.5 µM and substrate was held at 100-fold excess for all conditions. Reactions were conducted in 10 mM HEPES, pH 7.6. Values were determined at 60 minutes of reaction. Data are presented (in duplicate at each point) as a percentage of the highest assay value observed (set at 100) for each variable.

In the standard assay buffer, NoV Pro concentrations as low as 0.25 µM consistently produced signals at least 2-fold higher than background (no enzyme) control reactions (Fig. 1.3, top panel). Enzyme concentration of 0.5 µM and 1.0 µM were much superior to lower concentrations. Maximal efficiency of the reaction was observed at a 100-fold excess of substrate relative to NoV Pro (Fig. 1.3, bottom panel).
Figure 1.3. Effect of enzyme and substrate concentration on NoV Pro activity. Protease assays were conducted in standard assay buffer with the substrate Edans-EPDFHLQGPEDLAK-Dabcyl as described in the methods section. Data for triplicate reactions at each time point are presented. Comparisons of different enzyme concentrations (top panel) were performed at a constant enzyme:substrate ratio of 1:100. Comparisons of enzyme:substrate ratios (bottom panel) were conducted using 1.0 \( \mu \)M enzyme.

Cations affected the rate of reaction but not affinity for the peptide substrate.

To provide a basis for potential mechanisms responsible for observed differences in protease activity under various assay conditions, enzyme kinetics and parameters for NoV Pro were determined. Overall, \( K_m \) and \( k_{cat} \) for the GI and GII proteases were essentially identical under the standard assay conditions (Table 1.1).
Differences in the relative activity of the GI Pro under different assay conditions also appeared to be due to changes in $k_{cat}$ and not alterations of $K_m$ (Table 1.1, Fig. 1.4B). For the GI Pro, $k_{cat}$ was 3-fold greater in Tris buffer at pH 8.8 than at pH 7.6, resulting in a 3-fold increase in enzymatic efficiency. At pH 7.6, $k_{cat}$ was 2-fold higher in HEPES relative to Tris-HCl, for the GI Pro, resulting in more than a 2-fold increase in reaction efficiency. The reasons for these differences are currently unclear. The effects of pH may potentially be due to differences in overall charge of the GI and GII enzymes (GI pl=7.87, GII pl=8.74) and/or overall charge of the substrate (-3.2 at pH 7.6 vs. -4.0 at pH 8.8) (Protein Calculator\textsuperscript{TM}, v3.3, www.scripps.edu).

The effects of two cations at concentrations that reduced protease activity approximately 2-fold were also studied. In HEPES, pH 7.6, the presence of 100 mM NaCl or 10 mM MgCl\textsubscript{2} did not affect $K_m$ significantly, but reduced $k_{cat}$ and enzymatic efficiency significantly (Table 1.1, Fig. 1.4C).

<table>
<thead>
<tr>
<th>enzyme</th>
<th>$K_m$ (µM)</th>
<th>$k_{cat}$ (sec\textsuperscript{-1})</th>
<th>$k_{cat}/K_m$ (sec\textsuperscript{-1}µM\textsuperscript{-1})</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 mM HEPES, pH 7.6</td>
<td>GI</td>
<td>18.37 ± 0.83</td>
<td>0.0052 ± 0.0001</td>
</tr>
<tr>
<td>10 mM HEPES, pH 7.6</td>
<td>GII</td>
<td>23.11 ± 2.35</td>
<td>0.0053 ± 0.0002</td>
</tr>
<tr>
<td>10 mM Tris-HCl, pH 7.6</td>
<td>GI</td>
<td>20.33 ± 2.71</td>
<td>0.0023 ± 0.0001\textsuperscript{@}</td>
</tr>
<tr>
<td>10 mM Tris-HCl, pH 8.8</td>
<td>GI</td>
<td>21.05 ± 2.09</td>
<td>0.0074 ± 0.0002\textsuperscript{@,#}</td>
</tr>
<tr>
<td>10 mM HEPES, pH 7.6+100 mM NaCl</td>
<td>GI</td>
<td>26.65 ± 2.96</td>
<td>0.0036 ± 0.0002\textsuperscript{@}</td>
</tr>
<tr>
<td>10 mM HEPES, pH 7.6+10 mM MgCl\textsubscript{2}</td>
<td>GI</td>
<td>22.74 ± 1.71</td>
<td>0.0020 ± 0.0001\textsuperscript{@}</td>
</tr>
</tbody>
</table>

\textbf{Table 1.1. Kinetic analysis of NoV Pro activity.} Analysis was performed in triplicate. \textsuperscript{@}, $p<0.001$ (two-tailed t test, GraphPad\textsuperscript{TM}, GraphPad Software, Inc.) vs. 10mM HEPES, pH 7.6. \textsuperscript{#}, $p<0.001$ (two-tailed t test vs. 10mM Tris-HCl, pH 7.6). $K_m$ values were not statistically different ($p>0.05$).
Figure 1.4. Kinetics of NoV Pro reaction. Reactions using the standard assay peptide Edans-EPDFHLQGPEDLAK-Dabcyl (average values from triplicate analyses) are displayed for
A) GI NoV Pro and GII NoV Pro in standard assay buffer B) HEPES versus Tris-HCl at pH 7.6, 
and C) effect of NaCl and MgCl₂ on protease activity.

Chymostatin is a potent inhibitor of human NoV protease.

Chymostatin is a peptide produced by actinomycetes with known protease inhibitor 
characteristics. Chymostatin is a reversible inhibitor that shows selectivity for chymotryptase-
like serine proteases and inhibits the NoV protease (154, 165). When screening for novel 
protease inhibitors in a FRET-based assay it is crucial to have a potent positive control for
inhibition comparison. We performed a dose-response with the GII Pro to determine the IC$_{50}$ and IC$_{90}$ values for chymostatin. Figure 1.5 shows that chymostatin displayed an IC$_{50}$ value of 1.06 µM and an IC$_{90}$ value of 5.67 µM. By confirming the potency of chymostatin in vitro, moving forward with a high-throughput protease assay, chymostatin can be utilized as a positive control.

**Figure 1.5. Chymostatin inhibits NoV protease activity.** Reactions were performed in triplicate in standard assay buffer using standard assay peptide (Edans-EPDFHLQGPE DLAK-Dabcyl) at a final concentration of 50 µM. % Activity was determined after 1 hour with 100% as the average of triplicate reactions in the absence of inhibitor. Concentrations of chymostatin ranged from 0.05-100 µM. Data was plotted using a sigmoidal dose-response (variable slope) curve using GraphPad Prism™ Software.

**Discussion**

The systematic investigation of assay and enzyme storage conditions, which has not been previously reported for the human NoV protease, has resulted in one of the most robust human NoV protease assays reported to date (93, 94, 137, 154, 163). Importantly, the optimization of assay conditions in this report allows for the efficient conduct of studies with this important viral enzyme to be conducted at a more physiologic pH (7.2-7.6) than those previously used (8.0-9.0).
Most previous NoV Pro assays utilized conditions which we have demonstrated to be suboptimal or inhibitory. In previous assays, the inhibitory buffer components (i.e. NaCl) has necessitated the use of 5 to 20-fold more enzyme per reaction, and much smaller substrate to enzyme ratios (0.1-10), the latter of which we have shown limits the assay dynamic range. Most previous assays used reaction buffers containing 50 mM sodium phosphate (or Tris-HCl) with 100-150 mM NaCl, and enzyme stored in 10-30% glycerol in the presence of 100-150 mM NaCl in sodium phosphate buffer. We attribute the increased activity of the current assay to the (i) use of higher (50%) glycerol concentrations, HEPES, and the lack of cations for enzyme storage, (ii) the absence of cations in the reaction buffer, (iii) lower concentrations of buffering agent, and (iv) the use of HEPES instead of sodium phosphate or Tris-HCl as the buffering agent.

In our studies, cations, pH, and buffer composition affected the relative rate of reaction \( k_{cat} \), but not affinity for the peptide substrate \( K_m \), consistent with a pattern of non-competitive inhibition. The basis for the increased activity in the current investigations may be attributed to the significant improvement in \( K_m \). For example, \( k_{cat} \) values in our studies at pH 7.6 were comparable to those observed at pH 8.0 for a human GI Pro in a recent report (154), but \( K_m \) values in our studies were approximately 40-fold lower, resulting in a proportionately greater efficiency of reaction \( k_{cat}/K_m \). \( k_{cat} \) values in our studies at pH 7.6 were 2 to 6-fold less than those observed for a Southampton virus protease (a GI NoV) at pH 8.5 (94), but \( K_m \) values in our studies were 16 to 80-fold lower (94).

**Materials and Methods**

*Construction of the expression vectors.* A clone, pETSumoNVpro (strain Hu/GI.1/8FIIa/1968/USA, accession number JX023285), obtained from K.O. Chang (Kansas
State University, USA), was used as the source material for the GI protease. The NoV protease gene was amplified from pETsumoNVpro using the forward primer NdeNVp (5'-TCATCAAAGCTTCCATTTCTAGTGCGGTTCGCCCTC-3’), and the reverse primer HindNVp (5’-TCATCAAAGCTTCCATTTCTAGTGCGGTTCGCCCTC-3’). NdeNVp introduced an NdeI site and a hexahistidine tag into the 5’end of the amplicon. HindNVp introduced a stop codon after the NoV Pro reading frame, a HindIII site, and six additional nucleotides to the 3’end of the amplicon. To facilitate efficient restriction digestion of the 5’NdeI site, an additional six nucleotides were added to the 5’end by re-amplifying the amplicon with the primers NdeNVp2 (5’-CATCAACATATGCAACATCATCACCACACCACCACCACCACCACCAC-3’) and HindNVp. Both the pET32a vector and the full length NoV Pro amplicon were digested with NdeI and HindIII, purified, and ligated to form the pET32NVpro3C-1 expression construct.

The GII NoV Pro gene was cloned from the cDNA clone pGEMT-Easy-NV41 (strain Hu/GII.4/ Ast6139/01/Sp, accession number AJ583672), provided by F. Parra (Universidad de Oviedo, Spain) using the forward primer NV3C-F (5’-TGCATCAACATCATCAACATCGCCCAAGCATCTTGTC-3’) and the reverse primer NV3C-R (5’-GCAGAATTCTCTATTCAAGTGTAGCCTCCCCT-3’). The forward primer contained a hexahistidine tag along with a 5’-TG di-nucleotide that formed a start codon after ligation into the end-filled NdeI site in the vector. The pET-32a vector was digested with NdeI and filled in with Pfu DNA polymerase to form a blunt end, and subsequently digested with EcoRI. The cloned NoV Pro fragment was digested with EcoRI and ligated into the pET-32a vector to form the pET32NVpro3C-2 expression construct. For both NoV Pro constructs, DH5α E. coli were transformed and selected for with 100 µg/mL ampicillin.
Protein expression and purification. Expression and purification of NoV proteases was performed following previously described procedures (166). Briefly, *E. coli* BL21 Star (DE3) cells transformed with the individual protease expression constructs were grown at 37°C in 750 mL of Luria broth containing 100 µg/mL ampicillin. Protein expression was induced with IPTG. Cells were harvested, lysed by sonication, centrifuged, and the supernatant incubated with Talon metal affinity resin (Clontech Laboratories, Inc.). The resin was then washed, and the fusion protein was eluted with imidazole. Eluted fractions were checked for purity by SDS-PAGE and quantified by spectrophotometry.

Fractions of highest concentrations were pooled and dialyzed against storage buffer (10 mM HEPES, pH 7.6, 50% glycerol, 5 mM DTT), and aliquots were stored at either 4°C or -20°C. Enzyme yield was at least 20 mg/750 mL culture and was pure (>95%) of other detectable proteins following analysis in SDS-PAGE. Studies on the storage of NoV Pro demonstrated that the presence of 50% glycerol is required for long term stability (data not shown). Currently, enzyme activity upon storage is stable for at least 4 weeks at 4°C and at least 12 months at -20°C or -70°C.

NoV protease assay. The protease assay depends on the sequence-specific cleavage of the 14 amino acid FRET peptide substrate, Edans-EPDFHLQGPEDLAK-Dabcyl, between Q and G (Zeitler et al., 2006) by active NoV Pro and the subsequent detection of fluorescence. This substrate was chosen since the sequence represents that of the NoV NS2-3 cleavage junction and was previously demonstrated to be cleaved efficiently (93). The substrate was custom synthesized commercially to >95% purity and HPLC purified (New England peptide, Inc.). The standard assay reaction consists of 10 mM HEPES (pH 7.6), 0.1% CHAPS, 10 mM DTT, 30%
glycerol, 0.5-1.0 μM NoV Pro enzyme and 25-100 μM substrate peptide (50-100:1 relative to NoV Pro). On addition of the substrate, the reaction was incubated at 37°C, and fluorescence was measured at excitation and emission maxima of 360 and 460 nm respectively in a SPECTRAmax® GEMINI-EM Fluorescence microplate reader (Molecular Devices Inc.).

Analysis of enzyme parameters. Kinetic analysis of the enzyme was performed in 1X assay buffer (10 mM HEPES, pH: 7.6, 0.1% CHAPS, 10 mM DTT, 30% glycerol) using 1.0 μM enzyme and 3.9 to 125 μM substrate peptide. Fluorescence formed over 20 minutes was monitored at 1.0 min. intervals at 37°C. Measured fluorescence was converted to products formed in μM, using a standard curve of free Edans. $K_m$ and $k_{cat}$ were determined by non-linear regression analysis using Prism™ software (v5.0a, GraphPad Software Inc.).
Chapter II

Proteolysis of peptides representing ORF1 cleavage sites by Pro and ProPol

Reprinted in part with permission from the following manuscript:

In this chapter, we have characterized a human NoV ProPol with regard to pH, buffer conditions, and the presence of cations, and provide the first measurements of its enzyme kinetics using a previously described assay (155). Using synthetic FRET peptides representing the five NoV ORF1 polyprotein cleavage sites, we determined kinetic values \( K_m \) and \( k_{cat} \) for a human GII NoV Pro and ProPol to define the enzymatic parameters that govern ORF1 polyprotein processing, and provide a mechanistic explanation for the observed processing order of NoV ORF1. Our studies demonstrate that the two ‘early’ sites exhibit a much higher rate of enzyme turnover \( (k_{cat}) \) compared to the three ‘late’ sites. The two least efficient ‘late’ sites have the lowest enzyme turnover coupled with \( K_m \) values that are higher than the other three cleavage sites.

Results

NoV Pro and ProPol share similar pH profile and cation sensitivities.

Buffer and assay conditions for human NoV Pro have been previously optimized for enzymatic assays (154, 155), but no studies have characterized NoV ProPol with respect to similar parameters to optimize protease activity. We have purified recombinant GII.4 Pro and ProPol to >95% purity (Figure 2.1). The activity of GII NoV ProPol was examined over a range of pH and the presence of various cations, and enzyme kinetics were determined using a standard assay peptide that has been demonstrated to be efficiently processed (Table 2.1) (93). The ProPol construct used for these studies contained an E181A change at the internal protease cleavage site of ProPol which has been demonstrated to prevent its own self-cleavage upon purification and under assay conditions, while still retaining enzymatic activity (113, 127, 137, 167).
Figure 2.1. SDS-PAGE analysis of recombinant NoV Pro and ProPol. Proteins were electrophoresed in a 4-12% bis-tris NuPAGE® polyacrylamide gel (Life Technologies) and stained with Coomassie Blue dye. Lane M, molecular weight marker. Lane 1, purified His-Pro (20.2 kDa). Lane 2, purified His-ProPol (77.0 kDa).

Table 2.1. Peptides representing NoV ORF1 cleavage sites. The 14 amino acid FRET peptides (Dabacyl-14AA-Edans) used for kinetic studies are listed. Amino acid numbering is from N-terminus to C-terminus as suggested by Schechter and Berger (1967). Asterisks denote scissile bond. The sequence and origin of the standard assay substrate was previously described (93).
After 90 minutes of incubation, Pro and ProPol had comparable overall activities (Figure 2.2A). The kinetic parameters ($k_{cat}$ and $K_m$) of NoV Pro and ProPol with the standard assay peptide (Table 2.2, Figure 2.2B) were not significantly different ($p>0.05$, unpaired t test).

**Figure 2.2. NoV Pro and ProPol have similar activities in standard assay.** Protease assays were conducted in standard assay buffer using standard assay peptide (Edans-EPDFHLQGPEDLAK-Dabcyl) as described in the methods section. A) Activity of Pro and ProPol. NoV Pro and ProPol concentrations were 0.5 µM and substrate was held at 100-fold excess for all conditions. Data are presented (in triplicate at each point) as mean RFU adjusted for background. B) Comparison of Pro and ProPol kinetics. Reaction velocity (means of two independent experiments performed in triplicate) was plotted over a range of substrate concentration (3.9-125 µM, using 1.0 µM enzyme).
Table 2.2. Kinetics of NoV Pro and ProPol with ORF1 substrates. Peptides used in study are shown in Table 2.1. Reaction velocities used for non-linear regression were the means of at least two independent experiments done in triplicate. # indicates $K_m$ value was significantly different from ‘early’ site $K_m$ values ($p<0.01$) for NoV Pro or ProPol by one-way ANOVA with Bonferroni’s multiple comparisons test (excluding standard assay peptide). $ indicates kinetic parameters were significantly different between Pro and ProPol ($p<0.01$, two-tailed t test). All statistical tests were performed using GraphPad Prism™ Software.

Overall, Pro and ProPol exhibited similar profiles with respect to different assay components and conditions. NoV ProPol exhibited an optimum pH ranging from 8.6-9.0, and followed the same trend of higher activity with increasing pH as observed for Pro (Figure 2.3A). However, ProPol retained more activity at lower pH compared to NoV Pro (64% vs. 34% at pH 6.8) (Figure 2.3A), indicating the polymerase domain increased protease activity at lower pH.

Both enzymes exhibited similar inhibition profiles in the presence of NaCl, KCl (Figure 2.3B), MgCl₂, and MnCl₂ (Figure 2.3C). NoV Pro appeared to be less sensitive to the effects of CaCl₂ compared to ProPol (Figure 2.3D).

<table>
<thead>
<tr>
<th>Peptide</th>
<th>$K_m$ (µM)</th>
<th>$k_{cat}$ (1x10⁻⁴ s⁻¹)</th>
<th>$k_{cat}/K_m$ (M⁻¹s⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nterm/NTPase</td>
<td>17.28 ± 1.99</td>
<td>109.2 ± 4.04₃</td>
<td>632₅</td>
</tr>
<tr>
<td>NTPase/p20</td>
<td>17.81 ± 2.41</td>
<td>278.6 ± 12.26</td>
<td>1564</td>
</tr>
<tr>
<td>p20/VPg</td>
<td>62.42 ± 1.65₈</td>
<td>0.92 ± 0.01</td>
<td>1.5</td>
</tr>
<tr>
<td>VPG/Pro</td>
<td>48.89 ± 2.97₈</td>
<td>1.18 ± 0.21₅</td>
<td>2.4₅</td>
</tr>
<tr>
<td>Pro/Pol</td>
<td>18.02 ± 4.78</td>
<td>7.61 ± 0.66₅</td>
<td>42₅</td>
</tr>
<tr>
<td>Std. Assay Peptide</td>
<td>23.47 ± 4.84</td>
<td>34.1 ± 2.48</td>
<td>145</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Peptide</th>
<th>$K_m$ (µM)</th>
<th>$k_{cat}$ (1x10⁻⁴ s⁻¹)</th>
<th>$k_{cat}/K_m$ (M⁻¹s⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nterm/NTPase</td>
<td>16.44 ± 2.93</td>
<td>178.6 ± 10.99₅</td>
<td>1086₅</td>
</tr>
<tr>
<td>NTPase/p20</td>
<td>18.30 ± 1.71</td>
<td>251.5 ± 7.67</td>
<td>1374</td>
</tr>
<tr>
<td>p20/VPg</td>
<td>77.93 ± 11.1₈</td>
<td>0.95 ± 0.05</td>
<td>1.2</td>
</tr>
<tr>
<td>VPG/Pro</td>
<td>46.24 ± 8.00₈</td>
<td>2.00 ± 0.01₅</td>
<td>4.3₅</td>
</tr>
<tr>
<td>Pro/Pol</td>
<td>14.02 ± 3.95</td>
<td>23.4 ± 1.44₅</td>
<td>167₅</td>
</tr>
<tr>
<td>Std. Assay Peptide</td>
<td>22.64 ± 3.63</td>
<td>37.9 ± 2.12</td>
<td>167</td>
</tr>
</tbody>
</table>
Figure 2.3. Effect of pH and buffering agent on NoV Pro and ProPol activity. NoV Pro data for panels A-C have been previously described (155) and are included for comparison. NoV Pro and ProPol concentrations were 0.5 µM and standard assay substrate was held at 100-fold excess for all conditions. Values were determined at 60 minutes of reaction, and mean values for triplicate reactions are presented. Bars denote standard deviations. Data are presented as a percentage of the maximum value in each panel. A) pH curves for NoV Pro and ProPol. Assays were conducted in 50mM HEPES (pH 6.8-8.0) and 50 mM Tris-HCl (pH 8.0-9.0). Values for HEPES and Tris-HCl were normalized using pH 8.0 data. Panels B-D) Effects of various cations. Reactions were conducted in 10 mM HEPES, pH 7.6.

Processing of ORF1 cleavage site peptides by NoV Pro and ProPol.

The processing order of NoV ORF1 has been well established with the Nterm/NTPase and NTPase/p20 ‘early’ sites being cleaved first. The remaining ‘late’ sites (p20/VPg, VPg/Pro, and Pro/Pol) are not cleaved as efficiently (113, 116, 117). However, there has been no determination of the factors controlling the observed processing order. We hypothesized that processing order is primarily controlled by the binding affinity and/or rate of reaction of NoV
protease for each site. To test this, we designed 14 amino acid (P7-P7') FRET peptides representing each of the 5 ORF1 cleavage sites (Table 2.1). Peptides of this length were used since a previous report demonstrated binding efficiency was significantly reduced for peptides that did not include the P5 amino acid (94).

After 60 minutes under standard assay conditions using an enzyme:substrate ratio of 1:15, both of the ‘early’ site peptides (Nterm/NTPase, NTPase/p20) had reached 75% and 90% cleavage, respectively, for Pro, and 80% and 100% cleavage, respectively, for ProPol (Figure 2.4A,B). By contrast, cleavage of the ‘late’ site Pro/Pol peptide had reached only 15% and 30% completion by Pro and ProPol respectively (Figure 2.4A, B). The remaining two ‘late’ peptides, p20/VPg and VPg/Pro, did not reach more than 5% cleavage after 1 hour by either Pro or ProPol (Figure 2.4A, B). To determine if these reactions could be completed further, 20 hour incubations of the peptides were performed with an enzyme:substrate ratio of 1:2 with Pro or ProPol (Figure 2.4C). After 20 hours of incubation, the ‘early’ peptides displayed the same level of cleavage observed previously. The Pro/Pol peptide reached 54% cleavage with ProPol, but did not undergo further appreciable cleavage by Pro, remaining at 17% cleavage after overnight incubation. The remaining ‘late’ peptides did undergo further cleavage overnight by both Pro and ProPol resulting in 5-10% completion (Figure 2.4C). A previous study using synthetic substrates did not observe any processing of the p20/VPg, VPg/Pro, or Pro/Pol peptides by NoV Pro after an overnight incubation (137).
**Figure 2.4. ORF1 peptides are cleaved by both NoV Pro and ProPol.** All assays were conducted in triplicate at each point, using standard assay buffer and peptides shown in Table 1. For panels A & B, 1 µM and 5 µM enzyme was used for ‘early’ and ‘late’ peptides respectively but enzyme:substrate ratio was the same for all reactions (1:15). A) Processing of ORF1 peptides by NoV Pro. B) Processing of ORF1 peptides by NoV ProPol. C) 20 hour incubation of Pro and ProPol with ORF1 peptides using 1:2 enzyme:substrate (5 µM enzyme) ratio. Asterisks (*) indicate a significant difference (*p<0.01) or (**p<0.001) in percent completion between Pro and ProPol for ORF1 peptides determined by two-way ANOVA with Bonferroni's posttest. All statistical tests were performed using GraphPad Prism™ Software.
Despite a fairly rapid initial rate of cleavage (especially for ProPol), the Pro/Pol peptide did not show substantial additional processing after approximately 45 minutes. Supplementation with additional protease after 90 minutes of reaction resulted in only 6.8% additional cleavage after a further 1.5 hours of incubation (Figure 2.5). We hypothesized that product inhibition could be responsible for the observed cleavage inefficiency, as previously demonstrated for some HCV protease reactions (168, 169). However, the addition of peptides representing both of the Pro/Pol cleavage products, P7-P1 and P1’-P7’ (Table 2.1), to the reaction had no effect on NoV protease activity on the Pro/Pol peptide when added to the reaction assay at concentrations up to 500 µM (data not shown).

**Figure 2.5. Product inhibition is not responsible for cleavage of the Pro/Pol peptide not reaching completion.** Data are presented as mean relative fluorescence units (RFU) adjusted for background and represent an experiment performed in triplicate. All assays were conducted using standard assay buffer (see Methods) with 50 µM substrate. 5 µM ProPol enzyme was added to all reactions initially. At 90 minutes, the concentration of ProPol was doubled in the “Additional Enzyme” treatment (black triangles) to determine if uninhibited enzyme could cleave the remaining substrate in the reaction. Bars (if visible) denote standard errors.
Kinetic Analysis of ORF1 peptides.

The kinetic parameters for each peptide with NoV Pro and ProPol are shown in Table 2.2, and a comparative summary of overall enzymatic efficiency is displayed in Figure 2.6. For both NoV Pro and ProPol, the ‘early’ sites (Nterm/NTPase and NTPase/p20) were cleaved with the highest velocities. Of the ‘late’ sites, only the Pro/Pol peptide produced overall enzyme efficiency greater than 10 M⁻¹ s⁻¹ for both Pro and ProPol. All ORF1 ‘early’ site peptides, along with the Pro/Pol ‘late’ peptide, displayed the same binding affinity ($K_m$) for Pro and ProPol. The two remaining ‘late’ peptides, p20/VPg and VPg/Pro, displayed a marked increase in $K_m$ with a 2 to 3-fold increase for Pro and a 2 to 4-fold increase for ProPol.

![Figure 2.6](image)

**Figure 2.6. Relative cleavage efficiencies of NoV ORF1 peptides by Pro and ProPol.** Amino acids surrounding scissile bond of ‘early’ and ‘late’ cleavage sites are shown. Asterisk (*) denotes scissile bond at cleavage site. Fold change listed in $k_{cat}/K_m$ for NoV ORF1 peptides is relative to that observed for the most efficient site (NTPase/p20, shaded in black) for Pro and ProPol, respectively.

40
Enzyme turnover ($k_{cat}$) for the peptides representing ORF1 cleavage sites varied by orders of magnitude (Table 2.2). For both Pro and ProPol, the p20/VPg and VPg/Pro peptides showed 50-200 fold lower enzyme turnover compared to ‘early’ sites. NoV ProPol cleaved Nterm/NTPase more efficiently ($k_{cat}/K_m$) compared to Pro. For the ‘late’ sites, Pro and ProPol cleaved the Pro/Pol site with the highest efficiency but ProPol cleaved this site nearly 4-fold more efficiently compared to Pro. The lowest processing efficiencies observed for both enzymes were the p20/VPg and VPg/Pro peptides. The VPg/Pro peptide was processed with 650 and 320-fold lower efficiency compared to NTPase/p20 for NoV Pro and ProPol respectively (Figure 2.6). The p20/VPg peptide had over a 1000-fold decrease in efficiency for both Pro and ProPol compared to NTPase/p20 (Figure 2.6).

Protease activity on the two early site peptides was significantly greater than that observed for the standard peptide used for the analysis of assay conditions, reflected by 5 to 10-fold increases in $k_{cat}/K_m$ (Table 2.2). These differences were primarily due to increases in $k_{cat}$. It is surmised that these differences could be due to the fact that the standard assay peptide is a GI NoV Nterm/NTPase sequence (93). This peptide’s amino acid sequence (Table 2.1) was not observed to be present in any of the GII NoV sequences in GenBank we examined. GI and GII Pro have been reported to cleave this GI Nterm/NTPase peptide with similar efficiencies (154, 155). Similarly, we observed that GI Pro exhibits equivalent activity to GII pro on the GII Nterm/NTPase peptide ($K_m$ 18.6 µM, $k_{cat}$ 125 x $10^{-4}$ s$^{-1}$, $k_{cat}/K_m$ 675 M$^{-1}$s$^{-1}$). This indicates that the noted differences in protease activity between the GI and GII Nterm/NTPase peptides are likely to be due to differences in their primary amino acid sequences.
Competition of ‘early’ and ‘late’ peptides.

In the polyprotein, all cleavage sites are conceivably present at any given moment. To determine if peptides representing the ORF1 cleavage sites could influence one another, a competition experiment was designed. The enzyme turnover of the p20/VPg peptide was over 300-fold slower compared to the NTPase/p20 peptide for NoV Pro (Table 2.2, Figure 2.6). We hypothesized that the p20/VPg peptide could act as a competitive inhibitor since it binds the active site of the NoV proteases. A competition experiment was performed using p20/VPg peptide (inhibitor) and NTPase/p20 peptide (substrate) over a range of concentration (Figure 2.7). The ‘late’ p20/VPg peptide efficiently inhibited the cleavage of the ‘early’ peptide NTPase/p20 in a competitive manner with an inhibitor constant, \( K_i \), of 19.5 µM.

![Figure 2.7. p20/VPg peptide inhibits NoV Pro cleavage of NTPase/p20.](image)

Substrate velocity curves were generated with the independent variables [S] (NTPase/p20) and [I] (p20/VPg). A global curve fit was used to determine \( K_i \) for the p20/VPg peptide using GraphPad Prism™ Software. Assays were conducted in standard assay buffer and reaction velocity (average values from duplicate reactions) was plotted over a range of substrate [S] and inhibitor [I] concentrations in µM. Data is presented as a Lineweaver-Burk (double-reciprocal) plot.
Discussion

In this report, we provide evidence that human NoV protease efficiency ($k_{cat}/K_m$) at each individual ORF1 cleavage site is the primary determinant controlling viral ORF1 polyprotein processing order. Relative human NoV protease activity on peptides representing all five ORF1 polyprotein cleavage sites was correlated with the established processing order. Previous studies have determined the human NoV ORF1 polyprotein cleavage patterns, the di-amino acid cleavage sites, and the processing order for the viral ORF1 polyprotein, although there is some minor disagreement in the literature regarding cleavage order for some of the late cleavage sites (113, 116, 117, 137). The fact that several, long-lived, precursor proteins (e.g. ProPol) are observed during in vitro self-cleavage reactions complicates the interpretation of the final cleavage kinetics (113). The human NoV protease cleavage sites have been grouped into two categories. Two sites containing ‘Q-G’ at the cleavage point are cleaved initially followed by cleavage at three sites containing ‘E-G/A’. A previous report provided initial evidence for differential sensitivity of NoV Pro and ProPol for peptides representing different polyprotein cleavage sites, but did not provide a mechanistic explanation for the observations (137).

In this report, we demonstrate that the rate of reaction ($k_{cat}$) is the apparent primary factor controlling NoV protease enzymatic efficiency. Binding affinity ($K_m$) was not significantly different for the Nterm/NTPase, NTPase/p20, and Pro/Pol peptides for either NoV Pro or ProPol. However, $K_m$ values for the two least-efficiently processed ‘late’ peptides (p20/VPg and VPg/Pro) had $K_m$ values that were 2 to 4-fold higher, demonstrating an additional controlling contribution from this parameter for some sites. Since the binding affinities ($K_m$) were relatively similar between the ORF1 sites, this may indicate that the ORF1 sites adopt a similar
conformation that allow similar binding affinities. However, the primary amino acid sequence may play a more important role in enzyme turnover ($k_{cat}$) since these values varied by orders of magnitude. The fact that the p20/VPg peptide acted as a competitive inhibitor ($K_i=19.5$ µM) towards NoV Pro when mixed with the NTPase/p20 peptide indicated that the simultaneous presence of different cleavage sites in the polyprotein may constitute another potential mechanism for regulating protease activity during infection.

Previously, there has been no mechanistic explanation for the observed human NoV polyprotein processing order. It is generally surmised that viral polyprotein processing order could be controlled by either the relative accessibility of protease cleavage sites due to initial polyprotein conformation and changes in this conformation following sequential protease cleavages or the relative affinity/activity of the protease for each cleavage sequence. The conformation of the human NoV ORF1 polyprotein has been previously suggested to influence proteolytic processing (116, 117). However, another study demonstrated that mutation of each of the individual human NoV ORF1 polyprotein processing sites to inhibit protease cleavage did not inhibit the ability of the viral protease to process the remaining sites (113). This indicated the NoV ORF1 cleavage sites could be processed independently of one another. Similarly, in most cases, mutating individual poliovirus and hepatitis C virus cleavage sites does not prevent cleavage at additional sites (170, 171). Enzyme efficiency on peptides representing the cleavage sites of the PV P2-P3 polyprotein and HIV-1 Gag precursor have been shown to correlate with the known processing order (172-175). Over 200-fold differences in enzyme efficiency were observed between the 2B/2C and 3C/3D sites for poliovirus (172). In HIV, several hundred-fold differences in efficiency were observed between p2/NC and CA/p2, NC/p1 sites (176, 177). In
the HIV Gag protein, only the CA/p2 processing site has been demonstrated to be dependent on cleavage at another site (176).

These studies indicated that the primary amino acid sequence at each cleavage site was a controlling factor of NoV protease enzyme efficiency. The primary amino acid sequence surrounding cleavage sites has been shown to be a strong determinant of enzyme efficiency for poliovirus and hepatitis C virus (171, 172, 178-180). Changes of individual amino acids surrounding the NoV Pro/Pol scissile bond have been shown to affect protease activity (113). However, due to the structural differences between the peptides used in these studies and the NoV polyprotein, the absolute enzyme kinetics for processing the ORF1 polyprotein could differ if structural elements outside of P7-P7’ are involved.

Previously, no direct comparisons of the activity of human NoV Pro and ProPol with respect to assay buffer components have been reported. Here, we demonstrate that human NoV ProPol has essentially identical overall enzymatic activity and properties, shares a similar pH sensitivity profile, and is inhibited to nearly the same extent by NaCl, KCl, MgCl2, MnCl2, and CaCl2 compared to that previously described for NoV Pro (155). In addition, the current studies provide the first measurements of enzyme kinetics for a GII.4 ProPol.

We demonstrate that human NoV ProPol has equivalent, or superior, activity to Pro for all five peptides representing the human NoV ORF1 polyprotein cleavage sites. A similar observation has been made using a cell-based cleavage assay that found that ProPol was equivalent or superior to mature Pro in cleaving a ‘early’ site construct in a cellular environment (138). This data suggests that ProPol may be the dominant viral protease in infected cells. Both human NoV Pro and ProPol have been shown to be active forms of the viral protease that
process the viral ORF1 polyprotein (113, 116, 117, 137). Differential roles of Pro and ProPol in ORF1 processing have been proposed (113, 137), but the extent to which ProPol participates in the latter stages of human NoV polyprotein processing is somewhat unclear. In previous studies, human NoV ProPol has been shown to only cleave the two ‘early’ sites (113, 116, 117), and in one instance, the VPg/Pro ‘late’ site in trans (137). The feline calicivirus (FCV), a vesivirus, and human sapoviruses do not produce mature Pro or Pol in their lifecycles, but utilize ProPol for both polyprotein processing and RNA replication (4, 167). In human NoV and MNV, Pro/Pol is cleaved more efficiently in mammalian cells than in vitro (45, 46, 53), but it remains unclear the extent to which Pro and/or ProPol cleave the ORF1 polyprotein. In MNV, mature Pro has been shown to be dispensable for ORF1 polyprotein processing in vivo by mutating the Pro/Pol cleavage site and observing ORF1 proteolysis (181). However, infectious virus was not recovered and the mechanism of this inhibition was not determined as effects of the mutation on viral replication were not clearly described. Although apparently dispensable for intracellular viral polyprotein processing, mature viral protease and polymerase appear to be essential for successful NoV replication, but the precise roles for all three enzymes in NoV replication require further investigation.

We present in this study a comprehensive comparison of protease activities of GII human NoV Pro and ProPol with respect to various assay conditions and components. We analyzed the enzyme kinetics of Pro and ProPol using peptides representing all five cleavage sites of the NoV ORF1 polyprotein to provide an underlying mechanism for the established NoV polyprotein processing order. Relative enzymatic efficiency ($k_{cat}/K_m$) at each cleavage site was consistent with the previously established polyprotein processing order, indicating that this is the primary
controlling factor. Differences in enzyme efficiency were primarily due to differences in enzyme turnover ($k_{cat}$). Enzyme turnover varied by orders of magnitude between ‘early’ and ‘late’ peptides, and could be due to the stability of the enzyme:transition state complex. For example, the ‘early’ site peptides could have had stronger hydrogen bonding interactions with the protease that stabilized the enzyme:transition state complex while the ‘late’ peptides lacked such interactions. This could be a result of the primary amino acid sequence of the substrates. Higher $K_m$ values at the two most inefficiently processed sites were also observed. Overall, ProPol demonstrated equivalent or superior cleavage efficiency to Pro at all ORF1 cleavage sites, suggesting it may serve as the dominant form of the viral protease, similar to that observed for the vesiviruses and sapoviruses.

**Materials and Methods**

*Construction of the expression vectors.* GII human NoV *pro* and *propol* were cloned from the cDNA clone pGEMT-Easy-NV41 (provided by F. Parra, Universidad de Oviedo, Spain), GenBank accession AJ583672.2. The construction of the Pro expression vector was previously described (Viswanathan et al., 2013). NoV *propol* was amplified using the forward primer NV3C-F (5’-TGCATCACCATCACCATCACCACGCCCCCACCAAGCATCTGGTC-3’) and the reverse primer NVRp-R (5’-GCAGAATTCTCATTCGACGCCATCTTCATTCA-3’). The forward primer contained a hexahistidine tag along with a 5’-TG di-nucleotide that formed a start codon after ligation into the end-filled *NdeI* site in the pET-32a vector. The vector was digested with *NdeI*, filled in with *Pfu* DNA polymerase to form a blunt end, and subsequently digested with *EcoRI*. The cloned NoV *propol* fragment was digested with *EcoRI* and ligated into the pET-32a vector. The pET32NoVProPol construct was transformed into *E. coli* DH5α and selected for
with 100µg/ml ampicillin. Clones from resistant colonies were sequenced to ensure successful insertion and orientation.

To prevent self-cleavage of NoV ProPol during expression and purification, a Glu to Ala (E181A) mutation at the Pro/Pol cleavage site that retains protease activity (113) was engineered using QuikChange® Lightning Site-directed Mutagenesis Kit (Stratagene). The forward and reverse primers used for this mutagenesis were 5’-GAGGCTACACTTGCAAGTGAGCACAG-3’ and 5’-CTGTCACCACCTGCAAGTGAGCCTC-3’ respectively.

**NoV protease expression and purification.** Expression and purification of NoV Pro and ProPol were performed following previously described procedures (155, 166). Briefly, *E. coli* BL21 Star (DE3) cells were transformed with pET32NoVPro. To prevent the accumulation of early truncation products of ProPol, *E. coli* Rosetta™ (DE3) pLysS cells were transformed with pET32NoVProPol. Transformed cells were grown at 37°C in 750 mL of Luria broth containing 100 µg/mL ampicillin and 34 µg/mL chloramphenicol and transferred to fresh Luria broth containing 100 µg/mL ampicillin prior to induction. Protein expression was induced with 0.25 mM IPTG for 18 hours. Cells were harvested, lysed by sonication, centrifuged, and the supernatant incubated with Talon metal affinity resin (Clontech Laboratories, Inc.). The resin was then washed with 10 mM imidazole, and the fusion protein was eluted with 250 mM imidazole. Eluted fractions were checked for purity by SDS-PAGE and quantified spectrophotometrically. Fractions of highest concentrations were pooled and dialyzed against storage buffer (10 mM HEPES, pH 7.6, 50% glycerol, 10 mM DTT) (Figure 2.1). Aliquots were stored at -20°C.
**NoV protease assay.** Protease assays to characterize Pro and ProPol were performed as previously described (155). Briefly, the standard assay peptide (Table 2.1), an Edans-Dabcyl 14 amino acid FRET substrate (New England peptide, Inc.) was cleaved by active NoV Pro or ProPol enzyme in 1X assay buffer (10 mM HEPES, pH: 7.6, 0.1% CHAPS, 10 mM DTT, 30% glycerol) and fluorescence was measured at excitation and emission maxima of 360 and 460 nm respectively in a SPECTRAmax® GEMINI-EM Fluorescence microplate reader (Molecular Devices, Inc.).

**ORF1 peptide kinetic analysis.** Peptides (Edans-Dabcyl, 14 amino acids) were synthesized commercially to >95% purity and HPLC purified by New England Peptide, Inc. The peptides used were not ‘consensus’ sequences like the standard assay substrate previously described (93), but were designed to match the most frequent, naturally-occurring, amino acid sequences present in and around each of the five human NoV ORF1 polyprotein cleavage sites of the GII NoV entries filed in GenBank (http://www.ncbi.nlm.nih.gov/genbank). Peptides were designed by comparing 31 GII complete genome sequences. The most common P7-P7' sequence for each cleavage site was chosen (ranging from 52-84% prevalence). The amino acid sequence (P7-P7') surrounding the ORF1 cleavage sites from the 31 sequences examined are listed in Appendix I, Supplementary Tables 1-5.

Kinetic analysis of human NoV Pro and ProPol using peptides representing the human NoV ORF1 polyprotein cleavage sites (Table 2.1) was performed in 1X assay buffer (155) (10 mM HEPES, pH: 7.6, 0.1% CHAPS, 10 mM DTT, 30% glycerol). For these experiments, 1.0 μM enzyme was used for ‘early’ sites. Both 1.0 μM and 5.0 μM enzyme was used for the ‘late’ site peptides to compensate for lower protease activity. Two-fold serial dilutions of peptides
ranging from 3.9 to 125 µM were used for the Nterm/NTPase, NTPase/p20, and Pro/Pol peptides. For the p20/VPg and VPg/Pro peptides, 50, 100, 150, and 200 µM substrate was used to account for lower activity. Fluorescence formed over a 30 minute period of linear fluorescence signal was monitored at 1.0 min. intervals at 37°C for efficient peptides. Fluorescence signal was monitored at 5.0 min. intervals at 37°C for 75 minutes for the p20/VPg and VPg/Pro peptides to account for slower kinetics. Background-subtracted fluorescence (RFU) was converted to products formed in micromolar concentration using a standard curve of free Edans. Linear regression was used to determine reaction velocities for each concentration of substrate. \( K_m \) and \( k_{cat} \) were determined by non-linear regression analysis of the µM product/time slope values using GraphPad Prism™ Software.

**ORF1 peptide competition experiments.** Substrate-velocity curves were generated in the presence of various concentrations of the second (competitor) peptide. For these studies, the NTPase/p20 peptide was designated as the substrate, and the p20/VPg peptide was designated as the ‘inhibitor’ (I) or ‘competitor’. 1X assay buffer was used with 1.0 µM NoV Pro and two-fold serial dilutions of NTPase/p20 peptide ranging from 3.9 to 125 µM. A two-fold serial dilution of p20/VPg peptide ranging from 15.6-125 µM was used for competition. Background-subtracted, relative fluorescence units (RFU) from NTPase/p20 cleavage were converted into product formed in micromolar concentration. Nonlinear regression global curve fit was used to determine \( K_i \) for the p20/VPg peptide using GraphPad Prism™ Software.
Chapter III

Defining the core amino acid sequence controlling polyprotein processing order

Reprinted in part with permission from the following manuscript:

May J, Viswanathan P, Ng K.-S., Medvedev A, Korba B. 2014. The P4-P2' amino acids surrounding human norovirus polyprotein cleavage sites define the core sequence regulating self-processing order. J. Virol. Sep;88(18)10738-47. Copyright © American Society for Microbiology
We have previously demonstrated that human NoV polyprotein processing order is primarily an inherent property of enzymatic activity at each site, presumably due to the primary amino acid sequence surrounding NoV ORF1 cleavage sites (182). Altered self-processing order of the NoV ORF1 polyprotein has yet to be demonstrated. It remains unclear what comprises the minimal amino acid sequence surrounding ORF1 cleavage sites that controls enzyme efficiency and hence, polyprotein processing order. Based on published protease structural data, we hypothesized that the P4-P2’ residues of each ORF1 cleavage site should contain essentially all of the structural information necessary to control NoV Pro enzyme efficiency. In this report, using synthetic peptides (P7-P7’) and full-length ORF1 polyprotein in in vitro self-processing reactions, we have determined the core amino acid sequence controlling the processing kinetics of the NoV ORF1 cleavage sites. We also provide evidence that the P3 residue of the NS2-3 cleavage site participates in the control of NoV Pro enzyme efficiency, a role not previously appreciated in structural studies.

**Results**

*NoV protease enzyme efficiency is controlled by the P4-P2’ residues surrounding individual cleavage sites.*

The P4-P2’ residues of P7-P7’ synthetic FRET peptides representing the ‘early’ NS2-3 and ‘late’ NS6-7 cleavage junctions (Table 3.1, see Fig. 3.2 for schematic) were interchanged to determine if enzyme efficiency could be altered. A detailed description of the nomenclature used for the constructs containing amino acid substitutions is contained in the Methods section. Incubation of the wild-type NS2-3 (NS2-3 WT) peptide with NoV Pro displayed a much higher enzyme efficiency compared to the NS6-7 WT peptide (Fig. 3.1, Table 3.2) as previously
demonstrated (182). When the ‘early’ P4-P2’ residues of the NS2-3 site were changed to match the ‘late’ P4-P2’ of the NS6-7 cleavage site (NS2-3\textsubscript{LATE} peptide), cleavage kinetics were drastically reduced to essentially that observed for the NS6-7\textsubscript{WT} peptide (Fig. 3.1A). Insertion of the ‘early’ NS2-3 P4-P2’ residues into the ‘late’ NS6-7 cleavage junction in the NS6-7\textsubscript{EARLY} peptide dramatically enhanced cleavage kinetics to that closely resembling the NS2-3\textsubscript{WT} peptide (Fig. 3.1A).

<table>
<thead>
<tr>
<th>Peptide</th>
<th>P7</th>
<th>P6</th>
<th>P5</th>
<th>P4</th>
<th>P3</th>
<th>P2</th>
<th>P1</th>
<th>P1'</th>
<th>P2'</th>
<th>P3'</th>
<th>P4'</th>
<th>P5'</th>
<th>P6'</th>
<th>P7'</th>
</tr>
</thead>
<tbody>
<tr>
<td>NS2-3\textsubscript{WT}</td>
<td>L</td>
<td>G</td>
<td>D</td>
<td>Y</td>
<td>E</td>
<td>L</td>
<td>Q</td>
<td>*</td>
<td>G</td>
<td>P</td>
<td>E</td>
<td>D</td>
<td>L</td>
<td>A</td>
</tr>
<tr>
<td>NS2-3\textsubscript{LATE}</td>
<td>L</td>
<td>G</td>
<td>D</td>
<td>A</td>
<td>T</td>
<td>L</td>
<td>E</td>
<td>*</td>
<td>G</td>
<td>G</td>
<td>E</td>
<td>D</td>
<td>L</td>
<td>A</td>
</tr>
<tr>
<td>NS6-7\textsubscript{WT}</td>
<td>E</td>
<td>G</td>
<td>E</td>
<td>A</td>
<td>T</td>
<td>L</td>
<td>E</td>
<td>*</td>
<td>G</td>
<td>G</td>
<td>D</td>
<td>S</td>
<td>K</td>
<td>G</td>
</tr>
<tr>
<td>NS6-7\textsubscript{EARLY}</td>
<td>E</td>
<td>G</td>
<td>E</td>
<td>Y</td>
<td>E</td>
<td>L</td>
<td>Q</td>
<td>*</td>
<td>G</td>
<td>P</td>
<td>D</td>
<td>S</td>
<td>K</td>
<td>G</td>
</tr>
<tr>
<td>NS6-7\textsubscript{EARLY}P3-P2'</td>
<td>E</td>
<td>G</td>
<td>E</td>
<td>A</td>
<td>E</td>
<td>L</td>
<td>Q</td>
<td>*</td>
<td>G</td>
<td>P</td>
<td>D</td>
<td>S</td>
<td>K</td>
<td>G</td>
</tr>
<tr>
<td>NS6-7\textsubscript{EARLY}P4-P1'</td>
<td>E</td>
<td>G</td>
<td>E</td>
<td>Y</td>
<td>E</td>
<td>L</td>
<td>Q</td>
<td>*</td>
<td>G</td>
<td>G</td>
<td>D</td>
<td>S</td>
<td>K</td>
<td>G</td>
</tr>
<tr>
<td>GI NS2-3\textsubscript{WT}</td>
<td>E</td>
<td>P</td>
<td>D</td>
<td>F</td>
<td>H</td>
<td>L</td>
<td>Q</td>
<td>*</td>
<td>G</td>
<td>P</td>
<td>E</td>
<td>D</td>
<td>L</td>
<td>A</td>
</tr>
<tr>
<td>GI NS2-3\textsubscript{P3-Glu}</td>
<td>E</td>
<td>P</td>
<td>D</td>
<td>F</td>
<td>E</td>
<td>L</td>
<td>Q</td>
<td>*</td>
<td>G</td>
<td>P</td>
<td>E</td>
<td>D</td>
<td>L</td>
<td>A</td>
</tr>
</tbody>
</table>

Table 3.1. FRET peptides (Dabcyl-14AA-Edans) used in protease assays. Amino acids are numbered P7-P7’ from N-terminus to C-terminus using nomenclature described by Schecter and Berger (34). The GII “wild-type” (WT) sequences represent the most common, continuous P7-P7’ sequences observed in 75 GII.4 sequences analyzed from GenBank (31). The NS2-3\textsubscript{WT} sequence was found in 38/75 sequences examined. The NS6-7\textsubscript{WT} sequence was found in 46/75 sequences examined. The following GII.4 GenBank entries contained both NS2-3\textsubscript{WT} and NS6-7\textsubscript{WT} P7-P7’ sequences listed in Table 2: AGK25889, AGK25923, AGK25926, AFV08791, AGK25886, AGK25894, AGK25929, AFV08794, AFZ84662, ABO27559, AFP89595, AFL70022, AAR97662, and AAR97653. The GI NS2-3\textsubscript{WT} sequence was derived from accession number BAE43832. Bold letters indicate substitute variant amino acids that differ from the wild-type sequence for each respective site. Red amino acids indicate P4-P2’ residues unique to the GII WT NS2-3 site. Blue amino acids indicate P4-P2’ unique to the GII WT NS6-7 site. *Denotes cleavage site.
Figure 3.1. NoV Pro enzyme efficiency can be modulated by P4-P2’ substrate amino acid changes. Data are presented as mean relative fluorescence units (RFU) adjusted for background and represent two independent experiments performed in triplicate. All assays were conducted using standard assay buffer (see Methods) and 1 µM enzyme with 50 µM substrate. Bars denote standard errors. A) Effect of switching the P4-P2’ amino acids of WT NS2-3 and NS6-7. B) P4 and P2’ residues control enzyme efficiency. The P4-P2’, P3-P2’ and P4-P1’ amino acids of WT NS2-3 replaced the corresponding residues of WT NS6-7. Cleavage of WT NS2-3 from panel A is included for comparison (dashed line). C) P3-Glu enhances protease activity at the GI NS2-3 cleavage site.
Table 3.2 Enzyme efficiencies for GI and GII Pro with various peptide substrates. Reaction velocities used for non-linear regression to determine $K_m$, $k_{cat}$, and overall efficiency ($k_{cat}/K_m$) were the means of 2-4 independent experiments performed in triplicate. Standard errors are presented for $K_m$ and $k_{cat}$ values. All calculations were performed using GraphPad Prism™ Software. * Kinetics for WT peptides with GII Pro have been previously described (31) and have been included for comparison.

The NS2-3\textsubscript{LATE} peptide, which contained the NS6-7\textsubscript{WT} P4-P2’ sequence, had much lower enzyme efficiency ($k_{cat}/K_m$) compared to NS2-3\textsubscript{WT}, primarily due to a 20-fold decrease in $k_{cat}$ (Table 3.2). The overall enzyme efficiency of the NS2-3\textsubscript{LATE} peptide was comparable to that for the NS6-7\textsubscript{WT} peptide with $k_{cat}/K_m$ values of 40 and 42 M\textsuperscript{-1}s\textsuperscript{-1} respectively (Table 3.2). The NS6-7\textsubscript{EARLY} peptide, which contained the NS2-3\textsubscript{WT} P4-P2’ residues, displayed dramatically enhanced
kinetics compared to NS6-7WT with a greater than 20-fold increase in $k_{cat}/K_m$ (Table 3.2), again primarily due to an increase in $k_{cat}$ and comparable to that observed for the NS2-3WT peptide.

To confirm that the P4 residue was required for control of enzyme efficiency, an NS6-7 peptide that incorporated only the NS2-3 P3-P2’ ‘early’ residues was designed (NS6-7$_{EARYLP3-P2'}$). This peptide lacked the P4-Tyr of an ‘early’ site but instead retained the P4-Ala of the NS6-7WT ‘late’ cleavage site. The enzyme efficiency of the NS6-7$_{EARYLP3-P2'}$ peptide was essentially the same as that observed for the NS6-7WT peptide, approximately 30-fold lower than that observed for the NS6-7$_{EARYL}$ peptide (Table 3.2, Fig. 3.1B, A). These data demonstrated that P4-Tyr was critical for the high cleavage efficiency observed at ‘early’ sites.

To confirm that P2’ was critical for control of cleavage efficiency, a peptide was designed to incorporate only the NS2-3 P4-P1’ residues into the NS6-7WT peptide (NS6-7$_{EARYLP4-P1'}$). This peptide retained the NS6-7 P2’-Gly instead of the NS2-3 P2’-Pro. While the kinetics of the NS6-7$_{EARYL}$ peptide cleavage more closely resembled the kinetics of NS2-3WT (Table 3.2, Fig. 3.1B), the NS6-7$_{EARYLP4-P1'}$ peptide exhibited substantially enhanced cleavage efficiency compared to the NS6-7$_{EARYL}$ and NS2-3WT peptides (Table 3.2, Fig. 3.1B). Substitution of P2’-Gly for P2’-Pro increased $k_{cat}/K_m$ from 1169 to 3006 M$^{-1}$s$^{-1}$, demonstrating that the P2’-Pro of the NS2-3 site had a suppressive effect on enzyme efficiency and that P2’-Gly is more favorable for higher enzyme turnover ($k_{cat}$) and efficiency. These data demonstrated the importance of the P2’ residue for the NS2-3 site in the GII strain under study.

It is noteworthy that P2’-Pro is observed at the NS2-3, NS3-4 and NS5-6 cleavage sites across multiple GI and GII reference strains (Table 3.3), however further studies will be needed to determine if P2’-Pro suppresses enzyme efficiency across strains and genogroups as well as at
cleavage sites outside of NS2-3. Table 3.3 also illustrates a high degree of sequence conservation for the P4-P2’ amino acids between representative GI and GII strains.

<table>
<thead>
<tr>
<th>NS2-3</th>
<th>P4</th>
<th>P3</th>
<th>P2</th>
<th>P1</th>
<th>P1'</th>
<th>P2'</th>
</tr>
</thead>
<tbody>
<tr>
<td>GI Norwalk</td>
<td>F</td>
<td>H</td>
<td>L</td>
<td>Q</td>
<td>*</td>
<td>G</td>
</tr>
<tr>
<td>GI Chiba</td>
<td>M</td>
<td>H</td>
<td>L</td>
<td>Q</td>
<td>*</td>
<td>G</td>
</tr>
<tr>
<td>GI Southampton</td>
<td>F</td>
<td>H</td>
<td>L</td>
<td>Q</td>
<td>*</td>
<td>G</td>
</tr>
<tr>
<td>GII MD145</td>
<td>Y</td>
<td>E</td>
<td>L</td>
<td>Q</td>
<td>*</td>
<td>G</td>
</tr>
<tr>
<td>GII Camberwell</td>
<td>Y</td>
<td>E</td>
<td>L</td>
<td>Q</td>
<td>*</td>
<td>G</td>
</tr>
<tr>
<td>GII Hawaii</td>
<td>Y</td>
<td>E</td>
<td>L</td>
<td>Q</td>
<td>*</td>
<td>G</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>NS3-4</th>
<th>P4</th>
<th>P3</th>
<th>P2</th>
<th>P1</th>
<th>P1'</th>
<th>P2'</th>
</tr>
</thead>
<tbody>
<tr>
<td>GI Norwalk</td>
<td>F</td>
<td>Q</td>
<td>L</td>
<td>Q</td>
<td>*</td>
<td>G</td>
</tr>
<tr>
<td>GI Chiba</td>
<td>F</td>
<td>Q</td>
<td>M</td>
<td>Q</td>
<td>*</td>
<td>G</td>
</tr>
<tr>
<td>GI Southampton</td>
<td>F</td>
<td>Q</td>
<td>L</td>
<td>Q</td>
<td>*</td>
<td>G</td>
</tr>
<tr>
<td>GII MD145</td>
<td>Y</td>
<td>E</td>
<td>L</td>
<td>Q</td>
<td>*</td>
<td>G</td>
</tr>
<tr>
<td>GII Camberwell</td>
<td>Y</td>
<td>E</td>
<td>L</td>
<td>Q</td>
<td>*</td>
<td>G</td>
</tr>
<tr>
<td>GII Hawaii</td>
<td>Y</td>
<td>E</td>
<td>L</td>
<td>Q</td>
<td>*</td>
<td>G</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>NS4-5</th>
<th>P4</th>
<th>P3</th>
<th>P2</th>
<th>P1</th>
<th>P1'</th>
<th>P2'</th>
</tr>
</thead>
<tbody>
<tr>
<td>GI Norwalk</td>
<td>A</td>
<td>V</td>
<td>P</td>
<td>E</td>
<td>*</td>
<td>G</td>
</tr>
<tr>
<td>GI Chiba</td>
<td>A</td>
<td>T</td>
<td>S</td>
<td>E</td>
<td>*</td>
<td>G</td>
</tr>
<tr>
<td>GI Southampton</td>
<td>A</td>
<td>T</td>
<td>M</td>
<td>E</td>
<td>*</td>
<td>G</td>
</tr>
<tr>
<td>GII MD145</td>
<td>I</td>
<td>K</td>
<td>T</td>
<td>E</td>
<td>*</td>
<td>G</td>
</tr>
<tr>
<td>GII Camberwell</td>
<td>I</td>
<td>K</td>
<td>V</td>
<td>E</td>
<td>*</td>
<td>G</td>
</tr>
<tr>
<td>GII Hawaii</td>
<td>I</td>
<td>K</td>
<td>V</td>
<td>E</td>
<td>*</td>
<td>G</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>NS5-6</th>
<th>P4</th>
<th>P3</th>
<th>P2</th>
<th>P1</th>
<th>P1'</th>
<th>P2'</th>
</tr>
</thead>
<tbody>
<tr>
<td>GI Norwalk</td>
<td>I</td>
<td>N</td>
<td>F</td>
<td>E</td>
<td>*</td>
<td>A</td>
</tr>
<tr>
<td>GI Chiba</td>
<td>L</td>
<td>S</td>
<td>F</td>
<td>E</td>
<td>*</td>
<td>A</td>
</tr>
<tr>
<td>GI Southampton</td>
<td>I</td>
<td>S</td>
<td>F</td>
<td>E</td>
<td>*</td>
<td>A</td>
</tr>
<tr>
<td>GII MD145</td>
<td>L</td>
<td>S</td>
<td>F</td>
<td>E</td>
<td>*</td>
<td>A</td>
</tr>
<tr>
<td>GII Camberwell</td>
<td>L</td>
<td>S</td>
<td>F</td>
<td>E</td>
<td>*</td>
<td>A</td>
</tr>
<tr>
<td>GII Hawaii</td>
<td>L</td>
<td>S</td>
<td>F</td>
<td>E</td>
<td>*</td>
<td>A</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>NS6-7</th>
<th>P4</th>
<th>P3</th>
<th>P2</th>
<th>P1</th>
<th>P1'</th>
<th>P2'</th>
</tr>
</thead>
<tbody>
<tr>
<td>GI Norwalk</td>
<td>T</td>
<td>A</td>
<td>L</td>
<td>E</td>
<td>*</td>
<td>G</td>
</tr>
<tr>
<td>GI Chiba</td>
<td>T</td>
<td>T</td>
<td>L</td>
<td>E</td>
<td>*</td>
<td>G</td>
</tr>
<tr>
<td>GI Southampton</td>
<td>T</td>
<td>T</td>
<td>L</td>
<td>E</td>
<td>*</td>
<td>G</td>
</tr>
<tr>
<td>GII MD145</td>
<td>A</td>
<td>T</td>
<td>L</td>
<td>E</td>
<td>*</td>
<td>G</td>
</tr>
<tr>
<td>GII Camberwell</td>
<td>A</td>
<td>T</td>
<td>L</td>
<td>E</td>
<td>*</td>
<td>G</td>
</tr>
<tr>
<td>GII Hawaii</td>
<td>A</td>
<td>I</td>
<td>L</td>
<td>E</td>
<td>*</td>
<td>G</td>
</tr>
</tbody>
</table>

Table 3.3. Sequence alignment of P4-P2’ residues from ORF1 cleavage sites from GI and GII reference strains. The reference sequences used were GI Norwalk (NC_001959), GI Chiba (AB042808), GI Southampton (L07418), GII MD145 (AY032605), GII Camberwell (AF145896), and GII Hawaii (U07611).* Denotes scissile bond at cleavage site.
Self-processing order of NoV ORF1 polyprotein can be changed by altering P4-P2' residues.

To determine if the processing order of full-length ORF1 polyprotein could be changed, the P4-P2' amino acid changes described earlier were introduced into GI NoV ORF1 constructs placed downstream of a T7 promoter (Fig. 3.2). Due to our lack of a GII full-length ORF1 clone, GI ORF1 was used for these studies. We hypothesized the P4-P2’ residues would control polyprotein processing order in GI ORF1, expanding the FRET-based results using GII Pro and GII peptides. These constructs were used for in vitro transcription/translation reactions and the self-processing products were resolved by SDS-PAGE and autoradiography. The self-cleavage patterns of these constructs were followed over an eight hour time course (Fig. 3.3).

Figure 3.2. GI ORF1 constructs used for in vitro transcription/translation (TNT) reactions. GI NoV ORF1 (GenBank Accession M87661) is depicted with the predicted molecular weights for each mature protein. The P4-P2’ residues of the NS2-3, NS4-5, NS5-6 and NS6-7 cleavage sites are shown for each ORF1 construct. “WT” denotes wild-type sequence. Restriction sites used for construction of plasmids are shown.
Figure 3.3. Self-processing order of NoV ORF1 can be altered. $^{35}$S-Met labeled products were resolved by SDS-PAGE. ORF1 Pro$^{-}$ reactions, which lack a functional protease (13, 36) were used as negative controls. A) Early kinetics (15 minute intervals for 1 hour) of self-processing for ORF1WT, ORF1NS2-3LATE, ORF1NS6-7EARLY, and ORF1SWITCH constructs. B) Extended time course (1-8 hours) for constructs in panel A. C) 8-hour time course for ORF1NS4-5EARLY and ORF1NS5-6EARLY constructs. Arrows indicate key mature or precursor proteins.
Wild-type ORF1 (ORF1WT) (Fig. 3.3A, B, Lanes 1-4) displayed self-cleavage patterns similar to those previously reported (113, 116, 117). Two ‘early’ cleavage events (at sites NS2-3 and NS3-4) were observed that released mature Nterm and NTPase, which were prominent after only 30 minutes (Fig. 3.3A, Lane 2). The resulting p114 ‘late’ site precursor (p20VPgProPol) was present throughout the entire 8 hour time course (Fig. 3.3B, Lanes 1-4). Mature p20 was not detectable over the time course indicating the NS4-5 (p20/Vpg) junction was not processed efficiently. The ProPol precursor was faintly visible throughout the time course indicating low levels of processing at the NS5-6 (Vpg/Pro) site. Mature Pol was faintly detectable throughout the time course (Fig. 3.3B, Lanes 1-4). It was noted that, even at 15 minutes, no full length ORF1 was observed for the WT or other five ORF1 constructs (see lanes containing ORF1 Pro for comparison, which contains an inactive protease as described in the Methods section). However, an Nterm/NTPase (NS2-3) precursor was present for all constructs indicating that the ‘early’ site, NS3-4 (NTPase/p20) is likely processed co-translationally while the NS2-3 site is processed slightly less efficiently, requiring approximately 30 minutes for maximum cleavage. Consistent with this observation, we previously demonstrated that a peptide containing the NS3-4 cleavage site was processed more rapidly and displayed a 2-fold greater NoV Pro enzyme efficiency than a peptide containing the NS2-3 site (182).

The ORF1NS2-3LATE construct, which contained the NS6-7 P4-P2’ residues at its NS2-3 cleavage site (Figure 3.2), demonstrated altered kinetics that were consistent with the observations for the P7-P7’ FRET peptides described above. This P4-P2’ change was expected to slow processing down at the NS2-3 site, which was observed. The levels of mature Nterm and NTPase were significantly reduced over the time course and a dramatic increase in the 80 kDa
NtermNTPase precursor was evident (Fig. 3.3B, Lanes 5-8). ORF1\textsubscript{WT} had high levels of Nterm and NTPase after 30 minutes that did not substantially increase further over the 8 hour time course (Fig. 3.3B, Lanes 1-4). However, in the ORF1\textsubscript{NS2-3LATE} construct, mature Nterm or NTPase were not detectable until 60 minutes (Fig. 3.3A, Lane 9), and were present at drastically lower levels than observed for the WT polyprotein over the entire time course (Fig. 3.3B, Lanes 5-8).

The remaining cleavage sites appeared to be unaffected by the changes in the NS2-3 site as their processing matched that of ORF1\textsubscript{WT} (Fig. 3.3). In fact, for all constructs, only the sites containing sequence alterations displayed changes in processing efficiency (Fig. 3.3). This is consistent with previous studies that sought to alter processing kinetics at specific NoV ORF1 cleavage sites by changing single amino acids at or near the cleavage sites (17-19). In those studies, substitutions at one site that changed processing efficiency for that site did not affect processing kinetics at the other four (unaltered) sites (113, 116, 117).

In the ORF1\textsubscript{NS6-7EARLY} construct, the NS2-3 ‘early’ P4-P2’ residues replaced the P4-P2’ residues of the ‘late’ NS6-7 site (Fig. 3.2). These changes were expected to enhance the rate of cleavage between Pro and Pol relative to the WT polyprotein, which was observed. Within 45 minutes (Fig. 3.3A, Lane 12), mature Pol had accumulated to dramatically higher levels than observed for ORF1\textsubscript{WT} over 8 hours (Fig. 3.3B, Lane 4), indicating substantially increased processing at the NS6-7 site in the ORF1\textsubscript{NS6-7EARLY} polyprotein. In contrast to ORF1\textsubscript{WT}, the p114 (p20VPgProPol) precursor was present at substantially reduced levels (Fig. 3.3B, Lanes 10-13). However, the p20VPgPro precursor was prominent throughout the time-course for the ORF1\textsubscript{NS6-7EARLY} construct indicating the NS6-7 junction was cleaved very efficiently (Fig. 3.3B,
Lanes 10-13). Cleavage at the other sites was not affected by the changes made at the NS6-7 cleavage site.

In the ORF1\textsubscript{SWITCH} construct, the P4-P2' amino acids of the NS2-3 and NS6-7 sites were both interchanged (Fig. 3.2). Self-processing of this polyprotein displayed a combination of slowed processing at the NS2-3 site and more efficient cleavage at the NS6-7 site (Fig. 3.3B, Lanes 14-17). Cleavage at the other sites was not affected.

To determine if the P4-P2' core controlled ORF1 self-processing at the remaining two 'late' sites (NS4-5 and NS5-6), the 'early' P4-P2' sequence of NS2-3 replaced the P4-P2' core sequence at the NS4-5 (ORF1\textsubscript{NS4-5EARLY}), or the NS5-6 (ORF1\textsubscript{NS5-6EARLY}) cleavage sites in separate constructs (Fig. 3.2). As expected, the NS2-3 P4-P2' sequence enhanced the cleavage of the NS4-5 (p20/VPg) site in the ORF1\textsubscript{NS4-5EARLY} construct (Fig. 3.3C, Lanes 1-4). Mature p20 and the resulting precursor product (VPgProPol) were readily detected after only one hour indicating the NS4-5 site was cleaved very efficiently (Fig. 3.3C, Lanes 1-4). In ORF1\textsubscript{WT}, mature p20 was not detectable (Fig. 3.3B, Lanes 1-4).

The NS2-3 P4-P2' sequence also enhanced the cleavage of the NS5-6 (VPg/Pro) site. In the ORF1\textsubscript{NS5-6EARLY} construct, ProPol was present at one hour and accumulated to much higher levels compared to ORF1\textsubscript{WT} (compare Fig. 3.3C Lanes 6-9 vs. Fig. 3.3B Lanes 1-4). This indicated the NS5-6 site was cleaved very rapidly, releasing the ProPol precursor. Interestingly, p20VPg precursor (a product of NS5-6 cleavage) was not visible, possibly due to instability and would have to be determined in additional studies. In both the ORF1\textsubscript{NS4-5EARLY} and ORF1\textsubscript{NS5-6EARLY} constructs, the p114 (p20VPgProPol) precursor was not present to the extent of wild-type. This data demonstrated that all three 'late' sites, which are cleaved poorly in FRET based assays
as well as in in vitro translation assays (113, 116, 117), could be rapidly cleaved when the ‘early’ NS2-3 P4-P2’ core sequence replaced the natural P4-P2’ sequence. A summary of the GI NoV polyprotein self-processing results displayed in Figure 3.3 is summarized as follows:

- In WT, the ‘early’ NS2-3 and NS3-4 sites were cleaved rapidly
- In WT, the ‘late’ NS4-5, NS5-6, and NS6-7 sites were cleaved poorly resulting in the accumulation of precursor proteins.
- The ‘early’ NS2-3 P4-P2’ residues enhanced cleavage at the NS4-5, NS5-6, and NS6-7 cleavage sites after substitution.
- The ‘late’ NS6-7 P4-P2’ residues decreased cleavage efficiency at the NS2-3 cleavage site after substitution.

NS2-3 P3 residue is important for NoV Pro enzyme efficiency.

The P3 residues surrounding different NoV ORF1 cleavage sites display considerable variation, and have not been directly implicated in substrate recognition or have been shown to control enzyme efficiency (94-96, 107, 109, 110). However, our previous studies demonstrated a 4-fold difference in NoV Pro enzyme efficiency between GI and GII NS2-3 sites within synthetic peptides (182). Both GI and GII Pro displayed nearly identical kinetic profiles for GI and GII NS2-3 peptides (Table 3.2). This extends earlier findings where both GI and GII Pro displayed nearly identical kinetics on a GI NS2-3 substrate (154, 155). Both proteases cleaved the GII NS2-3WT peptide more efficiently compared to the GI NS2-3WT peptide (Table 3.3). For these peptides (see Table 3.1), a glutamate (Glu) at P3 was preferred over a histidine (His), indicating a possible role for P3 in interactions with the protease. However, these GI and GII NS2-3 peptides also differed at P4 (Table 3.1). At P4, the GI peptide contained a phenylalanine (Phe),
while the GII peptide contained a tyrosine (Tyr). For GII NoV ORF1, Phe and Tyr seem to be interchangeable (both are aromatic with similar side chains) at the NS2-3 P4 position since 48% of 75 GII.4 NS2-3 sequences examined from GenBank contain Phe while the remaining 52% contained Tyr (data not shown).

To directly determine the effects of a P3-Glu substitution, a GI NS2-3 peptide was designed to incorporate a P3-Glu instead of the wild-type P3-His (GI NS2-3P3-Glu) while retaining the same P4 residue as GI NS2-3WT (Table 2). The GI NS2-3P3-Glu peptide displayed higher enzyme efficiency (4-fold) compared to the GI NS2-3WT peptide (Fig. 3.1C, Table 3.2), demonstrating that P3-Glu was more favorable for higher enzyme efficiency compared to the P3-His. Comparable results were also obtained with GII Pro (Table 3.2) providing additional confirmation that composition of the substrate is the primary factor controlling NoV Pro enzyme efficiency.

**Discussion**

This report demonstrates that essentially all information necessary to control human NoV protease enzyme efficiency is contained within a minimal core sequence consisting of the P4-P2’ amino acids surrounding the viral ORF1 polyprotein cleavage sites. Using GII FRET peptides, we established the P4-P2’ residues regulate enzyme efficiency and extended these findings to a full-length GI ORF1 polyprotein. By interchanging the P4-P2’ residues of the NS2-3 and NS6-7 sites in the full-length ORF1 polyprotein, we demonstrated that this sequence is sufficient to transfer the relative protease efficiency of one donor site to a different site and alter NoV polyprotein self-processing order. We then verified these observations by enhancing self-processing at the NS4-5 and NS5-6 sites by substituting the P4-P2’ core from the NS2-3 ‘early’
site, demonstrating that protease processing at human NoV ORF1 cleavage sites is primarily regulated by the P4-P2’ sequences. These studies also confirmed our earlier observations that enzyme turnover is the primary factor regulating NoV protease efficiency (182).

Previous studies attempted to modulate NoV protease processing efficiency through single amino acid substitutions (113, 116, 117, 137). While providing valuable insight into the importance of individual amino acids, these studies did not effectively alter enzyme kinetics to fully switch the relative processing order of ‘early’ and ‘late’ sites. Due to the lack of a tissue culture system for the human NoV, it remains to be determined whether alterations in processing order will affect human NoV replication. However, in HCV, amino acid substitutions which enhanced cleavage at the NS4B/5A boundary were lethal; indicating the precursor protein involved must exist for a finite amount of time (146). Future studies in our laboratory will be aimed at introducing altered polyprotein processing patterns into a cultivatable Calicivirus.

As expected from previous crystallography-based studies (94-96, 107, 109), the P4 and P2’ amino acids were found to be critical components of the core amino acid sequence regulating NoV protease enzyme efficiency. Surprisingly, the P2’-Gly of the NS6-7 ‘late’ site was shown to be favored over the P2’-Pro of the NS2-3 ‘early’ site for enzyme efficiency. Changing the P4-P2’ amino acids of the NS6-7 site to match that of NS2-3 resulted in more comparable kinetics to WT NS2-3 than by changing P4-P1’, due to the suppressive effects of P2’-Pro. A similar observation has been made for the related Sapovirus where substitution of P2’-Asp for P2’-Pro at the NS2-3 site resulted in a 3-fold increase in enzyme efficiency, again highlighting the suppressive effects of P2’-Pro at the NS2-3 site (109). The restrictions on the main chain conformations available to Pro are unique and well known (183, 184). In contrast, Gly affords
the widest range of main chain conformations due to the lack of a side chain. Models indicate that main chain conformations with positive Φ torsion angles for the P2’ residue, which are only allowed for Gly, would allow the protease to form additional interactions with the portion of the peptide substrate C-terminal to the scissile bond (Fig. 3.4C, D). Additional kinetic studies and crystal structures with peptide substrates containing Gly at the P2’ position would assist in determining whether Gly-specific peptide substrate conformations help account for improvements in binding and activity.
Figure 3.4. Models of variations in the P3 and P2’ positions of peptide substrates. Complexes of NoV Pro GI protease (PDB: 2IPH) bound to four different substrates (containing P5-P4’ positions) were modeled. The structure of the protease is drawn as a semi-transparent solvent-exposed surface overlaid onto a stick representation. The structures of the substrates are displayed in stick representation, with carbon atoms colored magenta, oxygen red and nitrogen blue. The peptide sequences are shown below each figure and represent (A) GII NS2-3 cleavage site, (B) GI NS2-3 cleavage site, (C) GII NS2-3 cleavage site but with P5, P3’ and P4’ from the NS6-7 cleavage site, and (D) GII NS2-3 cleavage site but with P5, P2’-P4’ from the NS6-7 cleavage site. The figures were prepared using PyMOL (Schrödinger, LLC).
The importance of the P3 amino acid surrounding NoV cleavage sites in substrate recognition and enzyme efficiency has not been previously recognized in the literature. Our study provides the first empirical evidence to support a potential role for the P3 residue in modulating enzyme efficiency. We determined that substitution of a P3-Glu for the WT P3-His at the GI NS2-3 site increases NoV Pro enzyme efficiency 4-fold (Fig. 3.1C, Table 3.2), which provides an explanation for why a GII NS2-3 site (which contains P3-Glu) is cleaved more efficiently than GI NS2-3 by both GI and GII Pro (Table 3.3). While such interactions do not appear to be absolutely essential for some level of protease activity, it is clear that the composition of P3 can affect cleavage efficiency.

To better understand how the nature of the side chain at the P3 position modulates cleavage efficiency, we constructed models of human GI NoV Pro in the presence of peptides with variant P3 residues (Fig. 3.4A, B). The negatively charged side-chain carboxylate group of the glutamate P3 residue in the GII substrate is predicted to lie between the positively charged primary amino group of Lys-162 (GI Pro) or Arg-162 (GII Pro) and the amide nitrogen atom of the P1-Gln side chain. This model predicts that the favorable electrostatic and hydrogen-bonding interactions formed between the side chains of the P3-Glu, Lys-162 and P1-Gln residues promote the formation of a productive enzyme:transition-state complex, thus increasing \(k_{cat}/K_m\) and \(k_{cat}\). In contrast, these interactions are not observed in models of peptides with His at the P3 position, which is the case in the GI peptide substrate (Fig 3.4B). The imidazole side chain from a P3-His residue occupies a more shallow position in the P3 pocket due to the shorter and bulkier imidazole side chain. As a result, the histidine side chain is not predicted to interact with either
Lys-162 (GI) or Arg-162 (GII) (which are strongly conserved in the Norovirus genus) or the P1-Gln side chain. These models provide a possible explanation for why the enzyme:transition-state complex may be less stable in the GI peptide containing His at P3 when compared with the GII peptide containing Glu at P3.

These models also indicate that interactions at the P3 position may have been largely overlooked because most previous structural studies on human NoV protease activity, including our previous investigations (155, 182), utilized a Norwalk GI NS2-3 site (see Table 3.3) as a standard peptide substrate since this was one of the most efficient substrates in the founding literature and the crystalized human NoV enzyme was from a GI virus (22). As shown in Table 3.3, histidine appears to be predominant at the P3 of the NS2-3 site for GI NoV. Because our studies utilized GII NoV, where glutamate is predominant at the P3 of the NS2-3 cleavage site, these potential interactions have now become apparent.

Electrostatic interactions between charged side chains at the P3 position have previously been proposed to modulate the activity of the 3C-like processing protease from human coronavirus (185). A molecular model of a substrate peptide interacting with the S6-S5’ pockets in the coronavirus protease suggested that the positively charged side chain from a variant P3-Arg residue would form a favorable electrostatic interaction with the negatively charged side chain of Glu-166. This interaction was proposed to support the observation that the substitution of the wild-type P3-Val residue with the charged P3-Arg residue results in 5-14 fold increases in protease activity across multiple coronaviruses (185). Our observation of a similar effect in human NoV protease suggests that charged residues in the solvent-exposed P3 position may be a general strategy for modulating substrate binding and cleavage efficiency if charged residues are
present in the vicinity of the S3 pocket of the viral protease. This observation provides additional insight into understanding polyprotein processing and has important implications for improving the design of inhibitors.

**Materials and Methods**

The naming convention for the peptides utilized in these studies is as follows. Peptides are named for the NoV ORF1 cleavage site that they represent (e.g. NS2-3). Peptides that contained all “wild-type” sequences were designated with a ‘WT’ subscript (e.g. NS2-3WT). Peptides containing P4-P2’ amino acids substituted from an ‘early’ to a ‘late’ site, or *vice versa*, were denoted as either ‘early’ or ‘late’ based on whether ‘early’ or ‘late’ site P4-P2’ amino acids were used as the source (e.g. NS2-3LATE contains the P4-P2’ ‘late’ amino acids from the NS6-7 site, and NS6-7EARLY contains the P4-P2’ amino acids from the NS2-3 site). Peptides that contained amino acid changes other than P4-P2’ are indicated as such (e.g. the NS6-7EARLYP4-P1’ only incorporated the P4-P1’ residues of the NS2-3 site). Details of the composition of all peptides are contained in Table 3.1.

The naming convention for the ORF1 constructs followed an analogous format. The ORF1 construct that contained all WT sequences was named ORF1WT. When an individual cleavage site was switched from ‘early’ to ‘late’ or *vice versa*, the construct was denoted by which cleavage site was substituted and either ‘early’ or ‘late’ based on whether ‘early’ or ‘late’ site P4-P2’ amino acids were used as the source (e.g. ORF1NS2-3LATE contains P4-P2’ ‘late’ amino acids at the NS2-3 cleavage site). Details of the composition of all ORF1 constructs are contained in Figure 3.2.
**Kinetic analysis of NoV ORF1 peptides with NoV Pro.** Expression and purification of GI and GII NoV Pro has been previously described (155). The assay to monitor protease activity was adopted from earlier studies (154, 155, 182). Briefly, a 14 amino acid FRET peptide (Edans-Dabcyl, New England peptide, Inc.) was incubated with NoV Pro in 1X assay buffer (10 mM HEPES, pH:7.6, 0.1% CHAPS, 10 mM DTT, 30% glycerol) at 37°C. Fluorescence was then measured at excitation and emission wavelengths of 360 and 460 nm respectively using a SPECTRAmax® GEMINI-EM microplate reader (Molecular Devices, Inc.).

The GII wild-type (WT) peptides have been previously described (182) and represent the most prevalent P7-P7’ amino acid sequences surrounding the NS2-3 and NS6-7 cleavage sites of 75 GII.4 NoV entries filed in GenBank (http://www.ncbi.nlm.nih.gov/genbank). The GI WT NS2-3 peptide was adopted from an earlier study (93).

For kinetic analyses, 1.0 µM enzyme was used for the NS2-3WT, NS6-7EARLY, and NS6-7EARLYP4-P1’ peptides and 5.0 µM enzyme was used for the NS6-7WT, NS2-3LATE, and NS6-7EARLYP3-P2’ peptides to compensate for lower protease activity (Table 3.1). Kinetic parameters $K_m$ and $k_{cat}$ were determined for the peptides in Table 2 by methods previously described (155) using GraphPad Prism™ Software.

**Construction of NoV ORF1 plasmids for self-cleavage studies.** All ORF1 constructs (Fig. 2) were cloned from the previously described pNV-Neo (supplied by K.O. Chang, Kansas State University) (45). This GI sequence was derived from pNV101 (GenBank accession: M87661). The ‘wild-type’ ORF1 (ORF1WT) was amplified with NoVORF1-F and NoVORF1-R primers (Table 3.4) which placed ORF1 downstream of a T7 promoter and introduced unique XbaI and XhoI restriction sites (underlined) respectively. The ORF1 segment was subsequently ligated into
pET32a vector at the corresponding restriction sites and selected for with ampicillin (100 µg/mL). ORF1 Pro* contained a C1239A mutation of the WT ORF1 protease to prevent self-cleavage (118, 186) and was used as a negative control for in vitro transcription/translation reactions. ORF1 Pro* was constructed by site-directed mutagenesis using primers listed in Table 3.4.

Overlap-extension PCR was used to generate the ORF1NS2-3LATE and ORF1NS6-7EALRY constructs. For each site, two fragments were generated using either NS2-3 or NS6-7 1a + 1b and 2a + 2b primers (Table 3.4) which introduced a complementary 18 bp overhang that altered the P4-P2' coding sequence for both the NS2-3 and NS6-7 sites. These fragments were combined by overlapping PCR using primers 1a and 2b for each construct. For example, NS2-3-1a and NS2-3-2b fused 2 PCR fragments encompassing NS2-3 cleavage site and changed the P4-P2’ sequence. The altered NS2-3 fragment was digested with XbaI and NsiI and the altered NS6-7 fragment was digested with NsiI and NdeI before ligation into the digested WT ORF1 construct (Fig. 3.2). To make the ORF1SWITCH construct, a fragment encompassing the altered NS6-7 cleavage site in ORF1NS6-7EALRY was excised with NsiI and NdeI and ligated into the corresponding sites of the ORF1NS2-3LATE construct (Fig. 3.2).

The ORF1NS4-5EALRY and ORF1NS5-6EALRY constructs were generated by site directed mutagenesis of the ORF1WT construct using the Q5® Site-Directed Mutagenesis Kit (New England BioLabs Inc.) following the manufacturer’s protocol. Primers were designed using NEBaseChanger on-line tool (http://nebasechanger. neb.com) and are listed in Table 3.4.

In vitro transcription/translation (TNT) reactions. NoV ORF1 self-cleavage reactions were performed using TNT Quick-Coupled Transcription/Translation System (Promega). Each ORF1
construct (1.5 µg) was mixed with 20 µCi EasyTag $^{35}$S-methionine (PerkinElmer) in a 25 µL TNT reaction. Reactions were supplemented with 1.0 µL PCR enhancer (Promega) and incubated at 30°C. At each time point, 3.0 µL of each reaction was aliquotted and resolved on a 4-12% Bis-Tris NuPAGE polyacrylamide gel (Life Technologies), dried, and visualized by autoradiography.

<table>
<thead>
<tr>
<th>Primer</th>
<th>Oligonucleotide sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>WT ORF1 construction</strong></td>
<td></td>
</tr>
<tr>
<td>NoVORF1-F&lt;sup&gt;a&lt;/sup&gt;</td>
<td>GCATCTAGAGAAGGCGAGGCGAATGATGATGGCCTCAAGACG</td>
</tr>
<tr>
<td>NoVORF1-R&lt;sup&gt;a&lt;/sup&gt;</td>
<td>GCACTCGAGGGTATACGACGCATCATATTTACG</td>
</tr>
<tr>
<td><strong>ORF1 Pro construction</strong></td>
<td></td>
</tr>
<tr>
<td>GI NoVPro-Mut-F</td>
<td>CTTGCCACTATACCCGGAGACCGAGGACCACATACGTCCACAAG</td>
</tr>
<tr>
<td>GI NoVPro-Mut-R</td>
<td>CTTGTGGACGTATGGTGCCCCTGCTCCGGTAGCTGCGG</td>
</tr>
<tr>
<td><strong>Alt. NS2-3 construction</strong></td>
<td></td>
</tr>
<tr>
<td>NS2-3-1a</td>
<td>GAATTGAGCCGATAAACATTAC</td>
</tr>
<tr>
<td>NS2-3-1b&lt;sup&gt;b&lt;/sup&gt;</td>
<td>TCCACCTCTAGTCGCTCTGGCAGGAGTCAGTATAAG</td>
</tr>
<tr>
<td>NS2-3-2a&lt;sup&gt;b&lt;/sup&gt;</td>
<td>ACCGCCTAGAGGCTGGAGACCGACCACCTGGAAGCTG</td>
</tr>
<tr>
<td>NS2-3-2b</td>
<td>CTTCATATCTTCATCAGACCA</td>
</tr>
<tr>
<td><strong>Alt. NS6-7 construction</strong></td>
<td></td>
</tr>
<tr>
<td>NS6-7-1a</td>
<td>GACATGGTAGAAGGTTTGGCCTATG</td>
</tr>
<tr>
<td>NS6-7-1b&lt;sup&gt;c&lt;/sup&gt;</td>
<td>GGGGCCTCTGTAGATGAATTAGGGAATTGACCTCCAGGAGCTGAC</td>
</tr>
<tr>
<td>NS6-7-2a&lt;sup&gt;c&lt;/sup&gt;</td>
<td>TTCCACCTACAGGGCCGACAAGGGCAGTACCAGTGGTC</td>
</tr>
<tr>
<td>NS6-7-2b</td>
<td>GCCCCATATGAGACGCTTCGAC</td>
</tr>
<tr>
<td><strong>Alt. NS4-5 construction</strong></td>
<td></td>
</tr>
<tr>
<td>NS4-5 P4-P2′-Mut-F&lt;sup&gt;c&lt;/sup&gt;</td>
<td>CAGGGCCCACAAAAAGGAAGAAGACAAAAAG</td>
</tr>
<tr>
<td>NS4-5 P4-P2′-Mut-R&lt;sup&gt;c&lt;/sup&gt;</td>
<td>TAGATGGAAATCATTTGACCTACTATTTG</td>
</tr>
<tr>
<td><strong>Alt. NS5-6 construction</strong></td>
<td></td>
</tr>
<tr>
<td>NS5-6 P4-P2′-Mut-F&lt;sup&gt;c&lt;/sup&gt;</td>
<td>CAGGGCCCACACATGGAAGACCGAG</td>
</tr>
<tr>
<td>NS5-6 P4-P2′-Mut-R&lt;sup&gt;c&lt;/sup&gt;</td>
<td>TAGATGGAACTTTTACATTATAATCCACCTCTGTC</td>
</tr>
</tbody>
</table>

**Table 3.4. Primers used for in vitro transcription/translation constructs.** <sup>a</sup> Underlined sequences denote introduced restriction sites XbaI and XhoI for forward and reverse primers respectively. <sup>b</sup> Underlined sequences indicate the NS6-7 ‘late’ P4-P2′ coding sequence that was introduced at the NS2-3 cleavage site. <sup>c</sup> Underlined sequences indicate the NS2-3 ‘early’ P4-P2′ coding sequence that was introduced at the NS4-5, NS5-6, and NS6-7 cleavage sites.
3D Modeling of NoV Pro with GI/GII NS2-3 peptides. Peptide structures corresponding to cleavage sites lying at the junction of the Nterm (NS1/2) and NTPase (NS3) of both GI and GII NoV were modeled with reference to the extended reactive-site loop peptides observed in proteinase inhibitors bound to serine proteinases (187), and to the crystallographically determined structures of NoV GI protease bound to three types of peptides: (i) Michael acceptor covalent inhibitor (94), (ii) aldehyde covalent inhibitors (159), and (iii) natural and engineered C-terminal segments of NoV Pro mimicking the second product of the Pro/Pol cleavage reaction (96). All of these structures were determined at high resolution ($d_{\text{min}} = 1.5$-$2.4 \text{ Å}$) with clear definition of the bound peptides.

As a starting point for molecular modeling, the structure of NoV GI Pro bound to the Michael acceptor covalent inhibitor (2IPH) was modified by removing the covalently bound inhibitor and adjusting the side chains of the active site residues to form the hydrogen-bonding interactions expected for a catalytically active chymotrypsin-like protease. Because the 2IPH structure and other NoV Pro structures currently deposited in the PDB only provide information about peptides interacting with the S1-S4 binding pockets, the backbone conformations of the NoV GI and GII cleavage-site peptides were assumed to adopt conformations similar to those seen in 2IPH, as well as the canonical proteinase inhibitors and the uncleaved peptide bound to an inactive mutant form of the Foot-and-mouth disease virus 3C protease (111). Conformations of the side chains were modeled as low-energy rotamers that did not clash with the structure of NoV Pro. The complex structure was then refined by 100 steps of steepest descent minimization followed by 100 steps of conjugate gradient minimization using the Molecular Modeling Toolkit (MMTK) (188) as implemented in UCSF Chimera (189).
Chapter IV

Effects of altered human norovirus polyprotein processing on genome replication using a transient luciferase reporter replicon
A widespread strategy of positive-sense RNA viruses is to encode non-structural proteins in the context of a large polyprotein precursor. Tightly regulated processing patterns of viral polyproteins have been observed for HIV (190-193), alphaviruses (194, 195), HCV (180, 196, 197), and picornaviruses (198). As mentioned earlier, precursor intermediates in the proteolysis cascade can serve distinct functions in the viral lifecycle (e.g. precursor proteins may be the most favorable substrates for VPg nucleotidylation in picornaviruses (145)).

To date, few studies have focused on modulating or altering the processing order of viral polyproteins and determining their effects on replication. Due to the high conservation of polyprotein cleavage sites and regulation of processing efficiency, it would be expected that altered processing order would have detrimental effects on replication by interfering with the intricate timing of the processing cascade. We hypothesized that an intricately timed ORF1 processing pathway was essential for viral replication and altering the processing order of ORF1 cleavage sites would be detrimental.

In HCV, enhancing cleavage at the NS4B/5A junction inhibited viral replication highlighting the importance of inefficient cleavage at particular sites (146). In coxsackie B virus, decreasing processing at the 2B/2C junction abolished virus replication possibly due to failure to shut off host-cell translation (199). Also, substitutions that abolished processing at the poliovirus 2A/2B and 2C/3A junctions prevented virus recovery (170). These examples demonstrated that increasing or decreasing protease processing at polyprotein cleavage sites could have detrimental effects on virus replication.

In the related feline calicivirus (FCV), introducing an alanine at P1 for the NS2-3, NS3-4, NS4-5, and NS5-6 ORF1 cleavage sites was lethal for virus recovery, indicating it was critical to
fully process the polyprotein at these sites (75). Furthermore, in MNV, virus could not be recovered when cleavage was abolished at the NS6-7 junction (181).

Until recently, a reverse genetics system for human NoV was unavailable. Katayama et al., (2014) have generated a transient, luciferase-based, human NoV reporter replicon (46). As shown in Figure 4.1, this system requires transfection of a cDNA clone into Cos-7 cells that utilizes the eukaryotic elongation factor 1-alpha (EF-1α) promoter to drive expression of the GII.3 U201 human NoV genome which has a portion of ORF2 replaced with renilla luciferase (46, 200). Genome replication and subsequent subgenome expression results in luciferase expression and serves as the readout in this replicon system. Mutagenesis of the parental cDNA clone allows the effects of individual or multiple mutations on viral replication be determined.

Figure 4.1. Flowchart depicting human NoV transient replicon system. Luciferase-based reporter replicon developed by Katayama et al., (2014) is shown above.
We have previously demonstrated that the P4-P2’ amino acids surrounding the NoV ORF1 cleavage sites determine the efficiency with which a particular cleavage site is processed using in vitro translation assays (see chapter 3, (201)). In this study we sought to determine the effects of altered NoV ORF1 processing on genome replication using a transient human NoV replicon.

Results

Validation and optimization of NoV reporter replicon system.

We were kindly provided with the pU201-Rluc and pU201-Rluc-F/S plasmids (Figure 4.2) from Dr. Mary Estes (Baylor University) (46). To ensure the system was reproducible in our laboratory, we performed transfections utilizing 50 ng and 100 ng DNA plasmid per well of 96-well plate to determine which amount gave the best signal-to-background ratio (Figure 4.3). At 24 hours post transfection (hpt), mock samples had no detectable signal. The negative control, pU201-Rluc-F/S, which contains a deletion at base 4607 of the genome that introduced a frame-shift in the viral polymerase had measurements near $4 \times 10^5$ and $9 \times 10^5$ RLU for 50 ng and 100 ng DNA respectively (Figure 3). The WT plasmid, pU201-Rluc, gave measurements near $1 \times 10^6$ and $3.5 \times 10^6$ RLU for 50 ng and 100 ng respectively (Figure 4.3). At 24 hpt, the WT construct had statistically higher signal compared to the polymerase frame-shift mutant for both concentrations of DNA plasmid confirming the results reported by Katayama et al., (2014) (46). However, the 100 ng/well DNA yielded a higher signal-to-background ratio, so all additional experiments in this study used 100 ng/well DNA plasmid in a 96-well plate.

Of interest, we observed relatively high levels of luciferase expression in the polymerase frame-shift mutant (Figure 4.3). In theory, luciferase expression should only be observed as a
result of subgenomic RNA production since the luciferase gene has replaced VP1 of ORF2 in the NoV genome (Figure 4.2). Without a functional polymerase, we expected the polymerase-F/S mutant to have very low levels of luciferase expression. We hypothesized that translation termination/reinitiation between the polymerase and ORF2 could be the reason for such high background luciferase levels in our negative control. In support of this, a similar observation was made for a bovine NoV, where the VP1 capsid protein was made without subgenomic RNA production, but instead used a translation termination/reinitiation mechanism (202).

Figure 4.2. Constructs used for mutational studies using human NoV replicon. Site-directed mutagenesis was performed using pU201-Rluc (kindly provided by Dr. Estes, Baylor University) as template. Red residues indicate engineered mutations. The NS4-5 ‘LATE’ site was introduced into the NS2-3 and NS3-4 ‘EARLY’ sites. The NS3-4 “EARLY” site was introduced into the NS4-5, NS5-6, and NS6-7 ‘LATE’ sites. Alanine substitutions at the P1 residues were performed to abolish processing at specific cleavage sites (KO mutations).
Figure 4.3. Optimization of DNA transfection for human NoV replicon system. Varying amounts of DNA plasmid were used for DNA transfection with Lipofectamine 2000 to determine which amount gave the best signal-to-background separation. Data are presented in relative light units (RLU) as the average of quadruplicate wells 24 hours post transfection. Unpaired, two-tailed t tests were calculated using GraphPad (*** p<0.001).

To determine the time point at which viral replication yielded the highest signal, a time course was performed. Luciferase expression over the time course displayed a bell-shaped curve with the maximum signal being observed at 24 hpt (Figure 4.4). This coincides with the previous report by Katayama et al., (2014), which also observed maximum luciferase expression at 24 hpt (46). At 24 hpt, we observed 4.4-fold higher luciferase expression in the WT construct compared to the polymerase-F/S mutant (Figure 4.4). Similarly, Katayama et al., (2014), observed near 6-fold higher luciferase expression in WT versus the polymerase mutant at 24 hpt (46).
**Figure 4.4. Time course of human NoV replicon expression.** At each time point, cell lysates were collected and assayed for renilla luciferase. Data is a representative example of an experiment performed twice and is presented as the average RLU of transfected wells performed in triplicate at each time point. For all samples, 100 ng/well DNA plasmid was used for transfection. WT was compared to F/S control by two-way ANOVA with Bonferonni posttest to determine if luciferase expression was significantly higher in WT at each time point. *** \( p<0.001 \).

To confirm the viral protease was necessary for the replication of the luciferase-based replicon, a C139A mutation was engineered into the active site of NoV Pro (see Figure 4.2). Mutation of the active site nucleophile (cysteine to alanine) renders the viral protease inactive and polyprotein processing cannot occur. At 24 hours post-transfection, the pU201-Pro\(^\prime\) mutant did not replicate at levels above the F/S control (Figure 4.5) while WT replicated efficiently. This confirmed a functional viral protease is required for the replication of the NoV replicon and polyprotein processing is required to some extent.
Figure 4.5. NoV Pro activity is required for replication of NoV replicon. To prevent processing at all ORF1 cleavage sites, the nucleophile of the active site, Cys-139, was mutated to alanine using site-directed mutagenesis. After transfection into Cos-7 cells, luciferase expression was measured at 24 hours post transfection. Data is presented as the mean of 6 replicate samples. Error bars show standard error. A one-way ANOVA with Dunnett’s multiple comparison test was used to determine if means were statistically significant when compared to WT. *** \( p<0.001 \). \( pU201\)-Pro− was not statistically above the F/S control (\( p>0.05 \)).

**Effects of decreased processing at ‘early’ sites on genome replication.**

We have previously demonstrated that the NS2-3 and NS3-4 ‘early’ sites are cleaved very rapidly using both 14 amino acid peptides and full-length polyprotein translation assays (182, 201). Utilizing these same assays, we determined the NS4-5 site was the least efficient in regards to proteolytic processing. To determine the effects of decreasing processing at the NS2-3 and NS3-4 ‘early’ sites we substituted the P4-P2’ NS4-5 ‘late’ site (IKVE/GK) at both ‘early’ sites in the \( pU201\)-Rluc replicon (Figure 4.2).

At 24 hpt, the WT NoV reporter replicon, \( pU201\)-Rluc-WT, displayed luciferase expression levels 6-fold greater than the polymerase frame-shift control, \( pU201\)-Rluc-F/S.
(Figure 4.6). Both the NS2-3_LATE and NS3-4_LATE constructs displayed high levels of replication (Figure 4.6A). The NS2-3_LATE genome which contained P4-P2’ NS4-5 ‘late’ cleavage site residues showed a modest 30% decrease in luciferase expression (Figure 4.6A). The NS3-4_LATE genome which had NS4-5 P4-P2’ residues at its NS3-4 cleavage site did not show a decrease in luciferase expression (Figure 4.6A). These results indicated that decreased processing at the ‘early’ NS2-3 and NS3-4 cleavage sites did not inhibit genome replication.
Figure 4.6. Effects of switching ‘early’ and ‘late’ P4-P2’ cleavage sites on genome replication using human NoV replicon. Panel A. The NS2-3 and NS3-4 ‘early’ sites were replaced with the P4-P2’ residues of the NS4-5 ‘late’ site. Panel B. The NS4-5, NS5-6, and NS6-7 ‘late’ sites were replaced with the P4-P2’ residues of the NS3-4 ‘early’ site. Panel C. The second amino acid of the polymerase (G1194) was mutated and the NS6-7 site was changed from P4-P1’. Data is presented as the mean RLU at 24 hours post-transfection for an experiment performed in quadruplicate. All samples in the above figure were performed simultaneously in a single experiment and the Mock, F/S, WT, and NS6-7 EARLY data were transposed in different panels for comparison. Bars denote standard error. A one-way ANOVA with Dunnett’s multiple comparison test was used to determine if means were statistically significant when compared to WT. *** p<0.001, ** p<0.01, * p<0.05. ns indicates p>0.05 vs. pU201-F/S negative control.
**Effects of enhanced processing at ‘late’ sites on genome replication.**

We previously demonstrated using short peptides and full-length ORF1 polyprotein that the NS4-5, NS5-6, and NS6-7 ‘late’ sites are processed less efficiently by NoV Pro \textit{in vitro} (182, 201). We have also demonstrated that substitution of ‘early’ site P4-P2’ residues at any of the three ‘late’ sites will drastically increase proteolytic processing of a given ‘late’ site (201). The NS3-4 ‘early’ site is cleaved essentially co-translationally since the NTPase-p20 precursor cannot be detected in \textit{in vitro} translation assays. Thus, we used the NS3-4 P4-P2’ residues as the donor sequence to enhance processing at the NS4-5, NS5-6, and NS6-7 sites by substitution (Figure 4.2).

After transfection of the mutant constructs, luciferase expression was measured 24 hpt. Increasing processing at the NS4-5 and NS5-6 ‘late’ sites had little effect on genome replication and luciferase expression (Figure 4.6B). However, substitution of the P4-P2’ ‘early’ residues at the NS6-7 site resulted in a drastic reduction (75% decrease vs. WT) in replication and luciferase expression (Figure 4.6B). It was possible the mutations in the pU201-NS6-7\_EARLY were inhibiting the enzymatic activities of either Pro or Pol and the increased processing was not responsible for the substantial decrease in replication. For example, the P4-P1 mutations in Pro could have rendered the protease inactive, or the mutation at the P2’ (P1’ was unchanged) residue in the polymerase rendered the polymerase inactive (see Figure 4.2). Our earlier studies demonstrated mutation of the P4-P1 residues of the NS6-7 cleavage site had no effect on protease activity (201), thus we ruled out these mutations as a culprit for the decrease in replication of the NS6-7\_EARLY construct.
The second amino acid of the polymerase is important for genome replication.

To determine if the mutation (Gly to Pro) at P2’ in NS6-7 EARLY construct was responsible for the low replication ability of this construct, a point mutation at NS6-7 P2’ was introduced into the WT construct (pU201-G1194P). Unexpectedly, the point mutation at the second amino acid of the viral polymerase (Gly to Pro) significantly reduced (over 2-fold decrease) luciferase expression levels (Figure 4.6C). The second amino acid of the polymerase is largely unstructured and has not been implicated in serving a role in polymerase activity since mutation of this residue did not affect polymerase activity in vitro (83).

Since the G1194P mutation reduced replication over two-fold, we hypothesized the P2’ mutation was largely responsible for the drastic decrease in replication of the NS6-7 EARLY construct. To confirm, we constructed a P4-P1’ NS6-7 EARLY construct that would still increase processing at the NS6-7 site (See Chapter 3, Figure 3.1B), but would not include the mutation at the P2’ residue (Figure 4.2). We hypothesized the P4-P1’ mutation at the NS6-7 would ‘rescue’ replication by omitting the detrimental P2’ mutation and the increased processing at NS6-7 would have little effect on replication. Indeed, the NS6-7 EARLYP4-P1’ construct displayed luciferase expression levels comparable to WT (Figure 4.6C).

In conclusion, switching the P4-P2’ amino acids of ‘early’ and ‘late’ cleavage sites had little effect on genome replication. Decreasing processing at the efficient NS2-3 and NS3-4 cleavage sites had modest effects. Increasing processing at the inefficient NS4-5 and NS5-6 cleavage sites also had little effect. The P4-P2’ ‘early’ amino acids drastically reduced replication when inserted at the NS6-7 site. The P2’ Gly to Pro mutation at the NS6-7 cleavage site was responsible for the reduction in replication and was rescued when only the P4-P1’
residues were switched. This indicated increased processing at the NS6-7 site did not have any detrimental effects on replication. It remains unclear why mutation of the second amino acid of the polymerase has an inhibitory effect on replication.

Effects of preventing cleavage at individual ORF1 cleavage sites.

Since enhancing or reducing cleavage of ORF1 cleavage sites through P4-P2’ amino acid substitutions had modest effects on genome replication, we hypothesized that preventing cleavage of individual ORF1 cleavage sites would be necessary to see significant inhibition of genome replication. We sought to determine if abolishing cleavage at ORF1 sites would be detrimental as shown in the related FCV and MNV (75, 181). A series of mutations were constructed which introduced either Q→A or E→A mutations at the P1 positions of all ORF1 cleavage sites (see Figure 4.2). Alanine residues at the P1 positions of NoV cleavage sites prevent cleavage by NoV Pro (113, 181).

At 24 hours post-transfection, the effects of cleavage site “knockouts” (KO) on replication were examined. Surprisingly, all cleavage site knockouts produced luciferase expression well above the polymerase-F/S control (Figure 4.7). No effect on luciferase expression was observed for the ‘early’ site knockout mutants (NS2-3_KO and NS3-4_KO). This was unexpected since the ‘early’ sites are cleaved nearly co-translationally and we hypothesized this would be critical for genome replication. Next, the ‘late’ site knockouts (NS4-5_KO, NS5-6_KO, and NS6-7_KO) showed slight reductions in luciferase expression (Figure 4.7). Although we have speculated these ‘late’ site precursors could have distinct functions in replication, we were surprised that abolishing cleavage at these sites could still support efficient replication.
Figure 4.7. Effects of preventing cleavage at individual ORF1 cleavage sites. To prevent processing at ORF1 cleavage sites, the P1 amino acid of each ORF1 cleavage site was mutated to alanine using site-directed mutagenesis. After transfection into Cos-7 cells, luciferase expression was measured at 24 hours post transfection. The above figure is a representative experiment that was performed at least twice. Data is presented as the mean of samples performed in quadruplicate. Error bars show standard error. A one-way ANOVA with Dunnett’s multiple comparison test was used to determine if means were statistically significant when compared to WT. *** $p<0.001$, ** $p<0.01$.

Discussion

Until recently, the effects of individual mutations on human NoV replication could not be determined due to the lack of a reverse genetics system. The development of a luciferase-based human NoV transient replicon (46) was a significant step in NoV research by providing a method to determine the importance of mutations on replication. Our previous work has determined the core sequence of amino acids surrounding the human NoV ORF1 polyprotein

88
cleavage sites that control enzyme efficiency and subsequently processing order (182, 201). Switching the P4-P2’ amino acids between ‘early’ and ‘late’ cleavage sites effectively switches the ORF1 polyprotein processing order in vitro (201). We introduced analogous P4-P2’ amino acid changes into the GII.3 U201 luciferase-based NoV replicon. Surprisingly, enhancing or reducing cleavage at ORF1 polyprotein processing had little effect on genome replication.

Similarly, in the coronavirus MHV (Murine Hepatitis Virus), substituting the P4-P1 residues (decreases enzyme efficiency) of the cleavage site 3 (CS3) at the CS1 and CS2 cleavage sites had little effect on virus accumulation (203). In Rous Sarcoma Virus (RSV), enhancing proteolytic processing at the PR-NC and MA-p2 junctions as well as decreasing processing at the PR-NC junction had no effect on virus production (204). These examples highlight the ability of different viruses to accommodate alterations in processing efficiency at cleavage sites and retain efficient replication.

In the current study, we abolished cleavage at each individual ORF1 cleavage site by introduction of alanine at the P1 position and determined if genome replication could occur. Surprisingly, our data indicated genome replication could occur when each cleavage site was mutated individually. By virtue, this indicated genome replication could occur in the presence of NtermNTPase (NS2-3), NTPasep20 (NS3-4), p20VPg (NS4-5), VPgPro (NS5-6), and ProPol (NS6-7) precursors and no individual ORF1 protein was required to mature. This supported our earlier findings where ProPol was equivalent or superior to mature Pro in regards to protease activity (182). Furthermore, another group has demonstrated that VPgPro was the only form of the viral protease detected in B cells that supported NoV replication (41). These examples
provide evidence for a NoV replication scheme that is capable of accommodating precursor proteins as the sole source of enzymatic functions (i.e. proteolysis).

These results are in stark contrast to earlier studies in FCV, where alanine substitution at any ORF1 cleavage junction P1 prevented virus recovery (75). However in these earlier studies, intracellular viral genome replication was not monitored, so it remains possible that the genome was capable of replication, but the mutations affected packaging and/or release. Therefore, due to the nature of the human NoV replicon, where the capsid protein VP1 has been replaced with luciferase, preventing encapsidation and release, it is possible that the mutations preventing ORF1 processing imposed their inhibitory effect on packaging and release, and not genome replication.

In the process of switching the NS6-7 site to an ‘early’ site, we found that the second amino acid of the polymerase was critical for efficient replication. In WT, the second amino acid of the polymerase is glycine and when mutated to proline, a drastic reduction in replication was observed. This mutation was particularly interesting since the human NoV polymerase was crystallized with a glycine to serine mutation at this position (83). In these earlier studies, the Gly to Ser mutation at the second amino acid did not affect polymerase activity in in vitro assays as expected since the extreme N-terminus was not defined by electron density and is most likely disordered (83). It has been suggested that the polymerase Gly to Pro mutation may not affect mature Pol, but rather the precursor ProPol (Dr. Kenneth Ng, personal communication). Here, the Gly to Pro mutation may disrupt the function of ProPol, possibly through a conformational change. Proline can be a disruptive amino acid when substituted and may exert a negative effect on the conformation of ProPol since the second amino acid of Pol is in the linker region of the
precursor ProPol. Although enhancing processing at the NS6-7 junction had little effect on luciferase expression, it remains possible that ProPol existed for long enough to complete its role in replication but was not possible when the second amino acid of the polymerase was mutated.

In conclusion, switching ‘early’ and ‘late’ ORF1 cleavage sites and abolishing processing at cleavage sites had little effect on genome replication. Future studies will be aimed at confirming the presence of precursor proteins during replication in these mutant constructs. Also, analogous mutations will be introduced into a reporter virus system to determine if progeny virus can be recovered when individual ORF1 cleavage sites are blocked. Finally, further studies will be needed to determine the role of the second amino acid of the polymerase in maintaining efficient levels of replication.

Materials and Methods

Mutagenesis of parental pU201-Rluc (WT) construct. pU201-WT and pU201-F/S were kindly supplied by Dr. Mary Estes (Baylor College of Medicine). To generate the mutants shown in Figure 2, site-directed mutagenesis was performed using the Q5® Site-Directed Mutagenesis Kit (New England BioLabs Inc.) following the manufacturer’s protocol. Primers were designed and annealing temperatures were determined using NEBaseChanger on-line tool and are listed in Table 4.1.
Table 4.1. Primers used for site-directed mutagenesis of pU201-Rluc WT plasmid. Primers were designed using NEBasechanger on-line tool (New England Biolabs, Inc.).

<table>
<thead>
<tr>
<th>Primer</th>
<th>Oligonucleotide sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>pU201-Pro_F</td>
<td>ACCAGGTGACGCAAGGTGCCCCTAC</td>
</tr>
<tr>
<td>pU201-Pro_R</td>
<td>GTTGCTCAGGCTCAATG</td>
</tr>
<tr>
<td>pU201_NS2-3_LATE-F</td>
<td>GAAGGGAAGGAAGTCCGCCTGAAC</td>
</tr>
<tr>
<td>pU201_NS2-3_LATE-R</td>
<td>CACTTGATGTCACCGAGTAAGGGGGC</td>
</tr>
<tr>
<td>pU201_NS3-4_LATE-F</td>
<td>GAAGGGAAGGAATCTACAAACCTTCATTTTG</td>
</tr>
<tr>
<td>pU201_NS3-4_LATE-R</td>
<td>CACTTGATCTCATCATTCTCTCTCTG</td>
</tr>
<tr>
<td>pU201_NS4-5_EARLY-F</td>
<td>GAGGGCCCTAAGGCAAGACACATCC</td>
</tr>
<tr>
<td>pU201_NS4-5_EARLY-R</td>
<td>GAGCTCAGAATACCTGGATAGTCC</td>
</tr>
<tr>
<td>pU201_NS5-6_EARLY-F</td>
<td>CAGGGCCCTGAGCAAGGGCACC</td>
</tr>
<tr>
<td>pU201_NS5-6_EARLY-R</td>
<td>GAGCTCAGAATACCTGGATAGTCC</td>
</tr>
<tr>
<td>pU201_NS6-7_EARLY-F</td>
<td>CAGGGCCCTGAGCAAGGGCACC</td>
</tr>
<tr>
<td>pU201_NS6-7_EARLY-R</td>
<td>GAGCTCAGAATACCTGGATAGTCC</td>
</tr>
<tr>
<td>pU201_G1194P_F</td>
<td>TCTTGAGGGACCTGACGACAGGGGCACC</td>
</tr>
<tr>
<td>pU201_G1194P_R</td>
<td>GTGGTCTCACCCTGGCT</td>
</tr>
<tr>
<td>pU201_NS6-7_EARLY_2_3_F</td>
<td>CCAAGGCGGTAGCAACAGGGGACC</td>
</tr>
<tr>
<td>pU201_NS6-7_EARLY_2_3_R</td>
<td>AGCTCAGAATACCTGGATAGTCC</td>
</tr>
<tr>
<td>pU201_NS2-3_KO-F</td>
<td>CTACGAGATGGGCGGGCCAGAAGACC</td>
</tr>
<tr>
<td>pU201_NS2-3_KO-R</td>
<td>TCACGGAGTAAAGGGGGCA</td>
</tr>
<tr>
<td>pU201_NS3-4_KO-F</td>
<td>GTTCGAGCTGCCAGGCCCTAATC</td>
</tr>
<tr>
<td>pU201_NS3-4_KO-R</td>
<td>TCATCCATTCTCCTGTAAGG</td>
</tr>
<tr>
<td>pU201_NS4-5_KO-F</td>
<td>CATCAAAGTGGCAAGAAAGAGAG</td>
</tr>
<tr>
<td>pU201_NS4-5_KO-R</td>
<td>TCCTGGAATCGAGTATG</td>
</tr>
<tr>
<td>pU201_NS5-6_KO-F</td>
<td>GATAGATTTTGACGCCGCCCTAG</td>
</tr>
<tr>
<td>pU201_NS5-6_KO-R</td>
<td>TTCTCATTTGATGCTCAGACCTC</td>
</tr>
<tr>
<td>pU201_NS6-7_KO-F</td>
<td>GCCACCTCTTGCGGAGGTGAC</td>
</tr>
<tr>
<td>pU201_NS6-7_KO-R</td>
<td>TTCACTCTCTGCTCCTTG</td>
</tr>
</tbody>
</table>

Cell culture, transfections, and luciferase assays. Cos-7 cells (purchased from ATCC) were maintained in DMEM supplemented with 10% FBS, L-glutamine, and non-essential amino acids at 37°C + 5% CO2. Cos-7 cells were seeded in 96-well plates 24 hours prior to transfection.
at a density of 20,000 cells/well. Transfections were performed using Lipofectamine 2000 (Life Technologies) following the manufacturers protocol. At 24 hours post transfection, cells were washed with PBS, and lysed in 50 μL passive lysis buffer (Promega) for 15 minutes at room temperature. Next, 20 μL of lysate was assayed for renilla luciferase following the manufacturers protocol (Promega), using a Berthold Centro LB microplate luminometer.
CONCLUSION

The work presented in this dissertation provides an extensive examination of the mechanisms controlling human NoV polyprotein processing. The enzyme kinetics of the NoV protease are mainly controlled by the primary amino acid sequence surrounding each polyprotein cleavage site. Specifically, the P4-P2' core sequence contains the structural information to transfer the relative cleavage efficiency from one site to another. Also, the P3 residue surrounding cleavage sites may play a role in enhancing or reducing enzyme efficiency at a given site. Our work has the potential to improve the design of peptidomimetic protease inhibitors that are largely designed using the substrate specificity data available. The importance of polyprotein processing order on viral replication must be explored further. Our initial data indicates the absolute processing order of the NoV polyprotein may not be as critical as we originally hypothesized in regards to genome replication. Until a robust human NoV infection system becomes available, other cultivatable caliciviruses can be used to determine the effects of altered polyprotein processing on virus production.
**APPENDIX I**

| Most common | L | G | D | Y | E | L | Q | G | P | E | D | L | A | V | 67% prevalent |
|-------------|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Nterm/NTPase | L | G | D | Y | E | L | Q | G | P | E | D | L | A | V | Hu/NLV/GII/MrD145-12/1987/US |
| Nterm/NTPase | L | G | D | Y | E | L | Q | G | P | E | D | L | A | V | NLV/GII/Langen1061-2002/DE |
| Nterm/NTPase | L | G | D | Y | E | L | Q | G | P | E | D | L | A | V | Hu/GII.4/Armidale/NSW390/2008/AU |
| Nterm/NTPase | L | G | D | Y | E | L | Q | G | P | E | D | L | A | V | Hu/12/HS210/2010/USA |
| Nterm/NTPase | L | G | D | Y | E | L | Q | G | P | E | D | L | A | V | Hu/GII.4/New Orleans/2010/USA |
| Nterm/NTPase | L | G | D | F | E | L | Q | G | P | E | D | L | V | V | Hu/GII/Shanghai/SH2/2008/CHN |
| Nterm/NTPase | L | G | D | Y | E | L | Q | G | P | E | D | L | A | V | Hu/GII.4/HS191/2004/USA |
| Nterm/NTPase | L | G | D | Y | E | L | Q | G | P | E | D | L | A | V | Hu/GII.4/Armidale/NSW390/2008/AU |
| Nterm/NTPase | L | G | D | Y | E | L | Q | G | P | E | D | L | A | V | Hu/GII.4/New Orleans/2010/USA |

**Supplementary Table 1**: Shown above are the sequences used to select the most common GII.4 P7'-P7' sequence surrounding the Nterm/NTPase ORF1 cleavage site to be used for FRET peptide synthesis. The first row of the table shows the most common, continuous P7'-P7' sequence as well as the % prevalence. Amino acids shaded in gray represent residues that vary from the most common sequence.
Most common | L | D | E | F | E | L | Q | G | P | A | L | T | T | F |
---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
NTPase/p20 | L | D | E | F | E | L | Q | G | P | V | L | T | T | F | NLV/GII/Langen1061/2002/DE
NTPase/p20 | L | D | E | Y | E | L | Q | G | P | T | P | T | T | F | Hu/GII.12/5210/2010/USA
NTPase/p20 | L | D | E | F | E | L | Q | G | P | A | L | T | T | F | Hu/GII.4/New Orleans/2010/USA
NTPase/p20 | L | D | E | F | E | L | Q | G | P | A | L | T | T | F | Hu/GII/Shanghai/SH2/2008/CHN
NTPase/p20 | L | D | E | F | E | L | Q | G | P | A | L | T | T | F | Hu/GII-4/Ohio/7G/2012/USA
NTPase/p20 | L | D | E | F | E | L | Q | G | P | A | L | T | T | F | Hu/GII-4/Ohio/7I/2012/USA

**Supplementary Table 2:** Shown above are the sequences used to select the most common GII.4 P7-P7' sequence surrounding the NTPase/p20 ORF1 cleavage site to be used for FRET peptide synthesis. The first row of the table shows the most common, continuous P7-P7' sequence as well as the % prevalence. Amino acids shaded in gray represent residues that vary from the most common sequence.
**Supplementary Table 3**: Shown above are the sequences used to select the most common GII.4 P7-P7’ sequence surrounding the p20/VPg cleavage site to be used for FRET peptide synthesis. The first row of the table shows the most common, continuous P7-P7’ sequence as well as the % prevalence. Amino acids shaded in gray represent residues that vary from the most common sequence.
<table>
<thead>
<tr>
<th>Most common</th>
<th>N</th>
<th>E</th>
<th>K</th>
<th>L</th>
<th>S</th>
<th>F</th>
<th>E</th>
<th>A</th>
<th>P</th>
<th>P</th>
<th>S</th>
<th>I</th>
<th>W</th>
<th>S</th>
<th>77% prevalent</th>
</tr>
</thead>
<tbody>
<tr>
<td>VPg/Pro</td>
<td>N</td>
<td>E</td>
<td>K</td>
<td>L</td>
<td>S</td>
<td>F</td>
<td>E</td>
<td>A</td>
<td>P</td>
<td>P</td>
<td>S</td>
<td>I</td>
<td>W</td>
<td>S</td>
<td>Hu/GII.4/H5194/2009/US</td>
</tr>
<tr>
<td>VPg/Pro</td>
<td>N</td>
<td>E</td>
<td>K</td>
<td>L</td>
<td>S</td>
<td>F</td>
<td>E</td>
<td>A</td>
<td>P</td>
<td>P</td>
<td>S</td>
<td>I</td>
<td>W</td>
<td>S</td>
<td>NLV/GII/Langen1061/2002/DE</td>
</tr>
<tr>
<td>VPg/Pro</td>
<td>N</td>
<td>E</td>
<td>R</td>
<td>L</td>
<td>S</td>
<td>F</td>
<td>E</td>
<td>A</td>
<td>P</td>
<td>P</td>
<td>S</td>
<td>I</td>
<td>W</td>
<td>S</td>
<td>Hu/NLV/GII/MD145-12/1987/US</td>
</tr>
<tr>
<td>VPg/Pro</td>
<td>N</td>
<td>E</td>
<td>K</td>
<td>L</td>
<td>N</td>
<td>F</td>
<td>E</td>
<td>A</td>
<td>P</td>
<td>P</td>
<td>S</td>
<td>I</td>
<td>W</td>
<td>S</td>
<td>Hu/GII.4/Armidale/NLV390/2008/AU</td>
</tr>
<tr>
<td>VPg/Pro</td>
<td>N</td>
<td>E</td>
<td>K</td>
<td>L</td>
<td>S</td>
<td>F</td>
<td>E</td>
<td>A</td>
<td>P</td>
<td>P</td>
<td>S</td>
<td>I</td>
<td>W</td>
<td>S</td>
<td>Hu/GII.12/H5210/2010/USA</td>
</tr>
<tr>
<td>VPg/Pro</td>
<td>N</td>
<td>E</td>
<td>K</td>
<td>L</td>
<td>S</td>
<td>F</td>
<td>E</td>
<td>A</td>
<td>P</td>
<td>P</td>
<td>S</td>
<td>I</td>
<td>W</td>
<td>S</td>
<td>Hu/GII.4/New Orleans/2010/USA</td>
</tr>
<tr>
<td>VPg/Pro</td>
<td>N</td>
<td>E</td>
<td>K</td>
<td>L</td>
<td>S</td>
<td>F</td>
<td>E</td>
<td>A</td>
<td>P</td>
<td>P</td>
<td>S</td>
<td>I</td>
<td>W</td>
<td>S</td>
<td>Hu/GII/Shanghai/SH2/2008/CHN</td>
</tr>
<tr>
<td>VPg/Pro</td>
<td>N</td>
<td>E</td>
<td>K</td>
<td>L</td>
<td>S</td>
<td>F</td>
<td>E</td>
<td>A</td>
<td>P</td>
<td>P</td>
<td>S</td>
<td>I</td>
<td>W</td>
<td>S</td>
<td>Hu/GII.4/Toyama2/2007/JP</td>
</tr>
<tr>
<td>VPg/Pro</td>
<td>N</td>
<td>E</td>
<td>K</td>
<td>L</td>
<td>S</td>
<td>F</td>
<td>E</td>
<td>A</td>
<td>P</td>
<td>P</td>
<td>S</td>
<td>I</td>
<td>W</td>
<td>S</td>
<td>Hu/GII.4/NLV/GII/MD145-12/1987/US</td>
</tr>
<tr>
<td>VPg/Pro</td>
<td>N</td>
<td>E</td>
<td>K</td>
<td>L</td>
<td>S</td>
<td>F</td>
<td>E</td>
<td>A</td>
<td>P</td>
<td>P</td>
<td>S</td>
<td>I</td>
<td>W</td>
<td>S</td>
<td>Hu/GII.4/Armidale/NLV390/2008/AU</td>
</tr>
<tr>
<td>VPg/Pro</td>
<td>N</td>
<td>E</td>
<td>K</td>
<td>L</td>
<td>S</td>
<td>F</td>
<td>E</td>
<td>A</td>
<td>P</td>
<td>P</td>
<td>S</td>
<td>I</td>
<td>W</td>
<td>S</td>
<td>Hu/GII.4/Armidale/NLV390/2008/AU</td>
</tr>
</tbody>
</table>

**Supplementary Table 4:** Shown above are the sequences used to select the most common GII.4 P7-P7’ sequence surrounding the VPg/Pro ORF1 cleavage site to be used for FRET peptide synthesis. The first row of the table shows the most common, continuous P7-P7’ sequence as well as the % prevalence. Amino acids shaded in gray represent residues that vary from the most common sequence.
**Supplementary Table 5**: Shown above are the sequences used to select the most common GII.4 P7-P7’ sequence surrounding the Pro/Pol ORF1 cleavage site to be used for FRET peptide synthesis. The first row of the table shows the most common, continuous P7-P7’ sequence as well as the % prevalence. Amino acids shaded in gray represent residues that vary from the most common sequence.
APPENDIX II

Title: Serine Protease Mechanism and Specificity
Author: Lizbeth Hedstrom
Publication: Chemical Reviews
Publisher: American Chemical Society
Date: Dec 1, 2002
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.
This is a License Agreement between Jared May ("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: Jared May
Customer address: 2001 N. Adams St.
ARLINGTON, VA 22201

License number: 3592251077755
License date: Mar 18, 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: None
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: full article
Format: electronic

Are you the author of this Elsevier article?: Yes
Will you be translating?: No
Title of your thesis/dissertation: Mechanisms controlling human norovirus polyprotein processing

101
Expected completion date: May 2015
Estimated size (number of pages): 120
Elsevier VAT number: GB 494 6272 12
This is a License Agreement between Jared May ("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.

Elsevier Limited

The Boulevard, Langford Lane
Kidlington, Oxford, OX5 1GB, UK

Supplier

Registered Company Number 1982084
Customer name Jared May
Customer address 2001 N. Adams St.

ARLINGTON, VA 22201

License number 3592270228240
License date Mar 18, 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 full article
Format electronic
Are you the author of this Elsevier article? Yes
Will you be translating? No
<table>
<thead>
<tr>
<th>Title of your thesis/dissertation</th>
<th>Mechanisms controlling human norovirus polyprotein processing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expected completion date</td>
<td>May 2015</td>
</tr>
<tr>
<td>Estimated size (number of pages)</td>
<td>120</td>
</tr>
<tr>
<td>Elsevier VAT number</td>
<td>GB 494 6272 12</td>
</tr>
</tbody>
</table>
Title: The P4-P2′ Amino Acids Surrounding Human Norovirus Polyprotein Cleavage Sites Define the Core Sequence Regulating Self-Processing Order

Author: Jared May, Prasanth Viswanathan, Kenneth K.-S. Ng et al.

Publication: Journal of Virology
Publisher: American Society for Microbiology
Date: Sep 15, 2014
Copyright © 2014, American Society for Microbiology

Permissions Request
Authors in ASM journals retain the right to republish discrete portions of his/her article in any other publication (including print, CD-ROM, and other electronic formats) of which he or she is author or editor, provided that proper credit is given to the original ASM publication. ASM authors also retain the right to reuse the full article in his/her dissertation or thesis. For a full list of author rights, please see: http://journals.asm.org/site/misc/ASM_Author_Statement.xhtml

Copyright © 2015 Copyright Clearance Center, Inc. All Rights Reserved.


122. **Sasaki J, Ishikawa K, Taniguchi K.** 2012. 3CD, but not 3C, cleaves the VP1/2A site efficiently during Aichi virus polyprotein processing through interaction with 2A. Virus research 163:592-598.


147. **Kusov Y, Gauss-Muller V.** 1999. Improving proteolytic cleavage at the 3A/3B site of the hepatitis A virus polyprotein impairs processing and particle formation, and the impairment can be complemented in trans by 3AB and 3ABC. Journal of virology 73:9867-9878.


