Biochemically Investigating the Roles of Papain-like Proteases from +ssRNA Viruses in Replication and Innate Immune Suppression

By

Ankoor Patel

A Thesis submitted to the Faculty of Graduate Studies of
The University of Manitoba
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Microbiology
University of Manitoba
Winnipeg, Manitoba, Canada

Abstract

Positive-sense single-stranded RNA (+ssRNA) viruses are highly diverse in the way their genomes are packaged, polyproteins are arranged, and the hosts they infect. They affect many lifeforms, from bacteria to humans. Although viruses from different families vary in the complete panoply of proteins they produce in the infected host, some proteins are characteristic of most +ssRNA viruses, one being the papain-like protease (PLP). PLPs are cysteine proteases that are often essential in viral replication by acting to cleave their own polyprotein(s), which is essential in liberating the functional subdomains from within the larger polyprotein. The polyprotein(s) is a protein that is translated from a single open reading frame within the genome; yet encapsulates multiple individual protein domains that are fused together, which maximizes genomic coding capacity. The individual domains may be rendered inactive if they are not released from their precursor form. The primary function of PLPs is well understood and has been characterized extensively for many different +ssRNA viruses; however, PLPs have been shown to be more than just endopeptidases.

+ssRNA viruses have very small genomes and therefore utilize programmed ribosomal frameshifting (PRF), allowing for the translation of multiple proteins from one gene. Interestingly, as seen in Porcine reproductive and respiratory syndrome virus, one of if its PLPs, Nonstructural protein 1β, acts as a transactivator of PRF to produce two variant viral proteins that appear to be essential in immune suppression. Herein, nsp1β is shown to interact with a cellular protein and the viral genome at key residues to facilitate PRF. A biochemical and biophysical investigation illustrates how this complex forms and provides further insight into the novelty behind viral and cellular proteins acting in conjunction as transcriptional transactivators of PRF.

Plant affecting +ssRNA viruses have been studied for decades but the amount of research done on them is disparate in comparison to those that affect mammals. Here we show yet another role for plant-affecting viral PLPs, termed PRO, from genera *Marafivirus* and *Benyvirus* acting as deubiquitinases to potentially corrupt the ubiquitin-dependent proteasomal degradation system or tamper with cellular signaling. We have been able to determine that these proteases act as ubiquitin hydrolases and have specificity towards different ubiquitin substrates. We have also determined cleavage sites within the marafiviral polyprotein that had not been determined before. Crystal structures of PRO from the marafivirus type member *Maize rayado fino virus*, show differences within the enzyme in its native and substrate-bound forms. Protein-engineered ubiquitin-like inhibitors were designed against these enzymes to act as competitive inhibitors with hopes of imparting agricultural resilience. A similar strategy was used against a deubiquitinating PLP from Severe acute respiratory syndrome coronavirus 2, termed PL^{PRO}.

Acknowledgements

I firstly would like to recognize my supervisor, Dr. Brian Mark. He has been one of my biggest supporters throughout this degree and has always found healthy, positive ways to keep me motivated and continuing to thrive. His style of guidance has taught me autonomy, which has become one of my greatest assets as a scientist. He has taught me so much, personally, scientifically and professionally and for that I will always be thankful.

A big thank you to my committee members Dr. Joe O'Neil and Dr. Peter Pelka. Thank you for all the time you have spent reading over my far too lengthy reports, providing insightful discussion during meetings, the many reference letters, and the overall support towards my research and degree over the years.

I would not have been able to do all this without the support of my lab members, both past and present. It has been a pleasure to work in a healthy workplace and I am lucky enough to call many of my colleagues my friends. Ben, Grishma, Cole, Veronica, Grace, Kristen and Anu—thank you for everything. A deep respectful thank you to Dr. Judith Winogrodzki who was an immensely talented scientist and will be missed by many. I would also like to thank the undergraduate researchers that had a hand in this research, Jessica McBride, Vanessa Hoi, Oxana Katova and Nathan Glowach.

Thank you to the many collaborators to help enhance all my research endeavours. Dr. Eric Snijder, Dr. Sachdev Sidhu, Dr. Wei Zhang, Dr. Jacky Chung, Dr. Melvin Bolton, Dr. John Weiland, Dr. Emmely Treffers, Dr. Trushar Patel and Dr. Markus Meier.

I would also like to thank the funding agencies over the years that have helped fund this research and my stipends: NSERC, CIHR and Research Manitoba. Thank you to all of the

department members. All the support staff, colleagues (many of whom are friends now), instructors for TAships, and professors. It has been a pleasure.

Lastly, I would like to dedicate this body of work and this whole process to my loving family and partner. They have all been with me in my corner when times were tough and when I was losing steam. They all knew how to bring my spirits back up when I would lose faith and hope. I'm so thankful to all of you for your care, encouragement and support. This literally could not have been accomplished without you. Thank you, dad, mom, Kashmal and Ursula.

Table of contents

List of Figures

List of Tables

List of Copyrighted Materials

Chapter 2:

Patel, A., Treffers, E.E., Meier, M., Patel, T.R., Stetefeld, J., Snijder, E.J., Mark, B.L.. Molecular characterization of the RNA-protein complex directing -2/-1 programmed ribosomal frameshifting during arterivirus replicase expression. *J Biol Chem*. 2020. doi:10.1074/jbc.RA120.016105

Chapter 3:

Patel, A., McBride, J.A.M., Mark, B.L.. The endopeptidase of the maize-affecting Marafivirus type member maize rayado fino virus doubles as a deubiquitinase. *J Biol Chem*. 2021. doi: 10.1016/j.jbc.2021.100957.

List of Abbreviations

+ssRNA	Positive single-stranded RNA virus
PLP	Papain-like protease
PLCP	Papain-like cysteine protease
PRF	Programmed ribosomal frameshift(ing)
PRRSV	Porcine reproductive and respiratory syndrome virus
Nsp1β	Nonstructural protein 1β
Nsp	Nonstructural protein
Nsps	Nonstructural proteins
RdRp	RNA-dependent RNA polymerase
ORF	Open reading frame
OTU	Ovarian tumor
ER	Endoplasmic Reticulum
Nsp	Nonstructural protein
HC-PRO	Helper component protease
NIa	Nuclear inclusion protein a
RSS	RNA silencing suppressor
siRNA	Small interfering RNA
DCLs	Dicer-like proteins
RISC	RNA-induced silencing complex
vsiRNA	Viral small interfering RNA
HEN1	HUA Enhancer 1
AGO	Argonaute
FMDV	Foot-and-mouth disease virus
L^{pro}	Leader protease
eIF	Eukaryotic initiation factor
DUB	Deubiquitinase
sg mRNA	Subgenomic mRNA
pp1a	Polyprotein 1a
PAM	Porcine alveolar macrophage
pp1ab	Polyprotein 1ab
ZF	Zinc finger
HVR	Hypervariable region
TM	Transmembrane
EAV	Equine arteritis virus
Ub	Ubiquitin
ISG15	Interferon-stimulated gene 15
SRCR	Scavenger receptor cysteine-rich
HP-PRRSV	Highly pathogenic porcine reproductive and respiratory syndrome virus
MLV	Modified live virus
LDV	Lactate dehydrogenase-elevating virus
RBM	RNA-binding motif
SAP	SAF-A/B, Acinus, PIAS

IPTG	Isopropyl β-D-1 thiogalatopyranoside
SAF	Scaffold attachment factor
PIAS	Protein inactivator of activated STAT
STAT	Signal transducer and activator of transcription
SHFV	Simian hemorrhagic fever virus
TM	Transmembrane
WT	Wild type
Ab	Antibody
MS	Mass spectrometry
TRIS	Tris(Hydroxymethyl)aminomethane
qRT-PCR	Quantitative reverse transcription-PCR
PCBP2	Poly-(C) binding protein 2
Co-IP	Co-immunoprecipitation
PCBP1	Poly-(C) binding protein 1
KH	K-homology
EMSA	Electrophoretic mobility shift assay
IFN	Interferon
CP	Coat protein
TYMV	Turnip yellow mosaic virus
MRFV	Maize rayado fino virus
PRO	Protease
HEL	Helicase
MET	Methyltransferase
CAT	Chloramphenicol acetyltransferase
LUC	Luciferase
RK	Rat Kidney
HEK	Human embryonic kidney
AUC	Analytical ultracentrifugation
SAXS	Small-angle X-ray scattering
GST	Glutathione S-transferase
LB	Lysogeny broth
MES	2-(*N*-morpholino)ethanesulfonic acid
Ni-NTA	Nickel-nitrilotriacetic acid
SUMO	Small ubiquitin-like modifier
PMSF	Phenylmethylsunfonyl fluoride
ml	Milliliter
DSF	Differential scanning fluorimetry
L	Liter
h	Hour
mM	Millimolar
M	Molar
mg	Microgram
mm	Centimeter
DTT	Dithiothreitol
TBE	**T**ris(Hydroxymethyl)aminomethane-**b**orate-**E**thylenediamine tetraacetic acid

PAGE	Polyacrylamide gel electrophoresis
µCi	Micro Curie
ng	nanogram
SDS	Sodium dodecyl sulfate
PBS	Phosphate-buffered saline
Rpm	Revolutions per minute
NMR	Nuclear magnetic resonance
DLS	Dynamic light scattering
UbV	Ubiquitin variant
BlVS	Blackberry virus S
GSyV	Grapevine Syrah virus 1
OLV3	Olive latent virus 3
UPS	Ubiquitin proteasome system
MR	Molecular replacement
OBDV	Oat blue dwarf virus
PEG	Polyethylene glycol
Ub-AMC	Ub-7-amido-4-methylcoumarin
Ubi-1	Ubiquitin 1
BNYVV	Beet yellow vein virus
BSBMV	Beet soil-borne mosaic virus
SARS-CoV-2	Severe acute respiratory syndrome coronavirus-2
PL^{pro}	Papain-like protease
Cryo-EM	Cryo-electron microscopy

1 Chapter 1: Literature Review

1.1 *Introduction to +ssRNA viruses*

Positive-sense single-stranded RNA (+ssRNA) viruses, or Group IV viruses by the Baltimore classification, are remarkably simple in their genomic architecture yet have an immense host range and have had devastating impacts on virtually all lifeforms. +ssRNA viruses are often thought of as being relics from a prebiotic world, presumably due to the RNA world hypothesis in that the earliest forms of life depended sheerly on RNA for the storage of genetic content and also for ribozyme-mediated catalysis events (1, 2).

The genomes of +ssRNA viruses act as mRNA and are directly accessible by the host cell ribosome. On average, +ssRNA viruses have the smallest genomes and also have the most restricted genome sizes, only varying by ~1 order of magnitude at the extremes in terms of kbp (3). Bacteria and fungi-affecting affecting +ssRNA viruses have a genome size of only ~4 kbp whereas algae, plant, invertebrate and vertebrate-affecting +ssRNA viruses predominantly range from 6-12 kbp, though coronaviruses can have genomes >30 kb (3, 4). Interestingly, for being so simple at the genomic level, they proportionally affect more eukaryotes than their DNA counterparts, yet they have no known archaea-affecting members (5). Their variation in host range and preference for eukaryotic systems suggest that their minimalistic genomes are not to be conflated with restrictive host choice. +ssRNA viral genomes code for their own RNA replication machinery, including the RNA-dependent RNA polymerase (RdRp), which imparts a large evolutionary benefit. The error-prone nature of RNA viral replication events dictated by the low fidelity of the RdRp, coupled with the rapid ability of genome replication due to small genome size manifests in the generation of a multitude of quasispecies that can often benefit the virus in

infectivity, immune evasion and replication (6, 7). These abilities also allow for host-switching if the mutations are drastic enough, which can clearly have potential benefits to the virus (6, 8).

Having such small genomes, +ssRNA viruses must be creative to produce all the proteins they need to successfully replicate in the host cell. This means they must be efficient with minimal genetic material and thus resort to polycistronic mechanisms to maximize coding capacity. The stability and topology of the RNA genome also makes it essential for these viruses to optimize their genomic coding capacity for RNA integrity and genome packaging (9). The translation of polyproteins and programmed ribosomal frameshifting (PRF) are two staples of +ssRNA viruses and aid in protein production through a disguised means and importantly do not require additional genetic elements to direct translation, again limiting the amount of genetic material needed in their genome (10). Interestingly many viral proteins are multifunctional and the papain-like cysteine proteases of certain +ssRNA viruses have been found to assist with both polyprotein processing, immune system antagonism, and programmed ribosomal frameshifting (4, 10)

1.2 *Viral papain-like cysteine proteases*

Papain is a cysteine protease found in the latex of the papaya fruit (11, 12). Originally described in 1971, the overall fold and amino acid sequence architecture has been used extensively to describe many different enzymes today, including viral proteases. Papain-like cysteine proteases (PLCPs or PLPs) are composed of two (or three) core domains with active sites containing a catalytic cysteine and histidine in close proximity, which are typically formed at the interface of two of the core domains as shown in Fig. 1.1A & B (12). Papain-like folds are typically divided into two domains in which the first is mainly composed of helices and contains the catalytic Cys (R domain), while the second domain is β-sheet rich and provides the catalytic His (L domain)

(11–13). Sequentially, their active sites characteristically have an N-terminally located Cys residue and a more C-terminally located His, and often a third polar residue that participates in the formation of a catalytic triad and found C-terminally to the His (12, 14–16).

PLCPs catalyze the hydrolysis of peptide (or iso-peptide) bonds within the proteins and have been shown to have specificity towards amino acids with smaller, less bulky side chains (14, 17). Within the active site of these enzymes, the imidazole group of the catalytic His side chain acts to decrease the pKa of the active site Cys, which results in deprotonation of the Cys thiol group and generation of the thiolate nucleophile (Fig. 1.2) (14, 16). The third polar residue of the active site (if present) acts to assist via orienting and polarizing the His imidazole group for more efficient thiol deprotonation (14, 16). Nucleophilic attack is carried out by the thiolate on the carbonyl carbon of the scissile peptide bond, resulting in the cleavage of the amino group, formation of a thioester intermediate and restoration of the deprotonated state of the His base (14, 16). The catalytic cycle is completed upon hydrolysis by a water molecule of the thioester bond, forming the carboxy terminus of the original C-terminal portion of the substrate (14, 16).

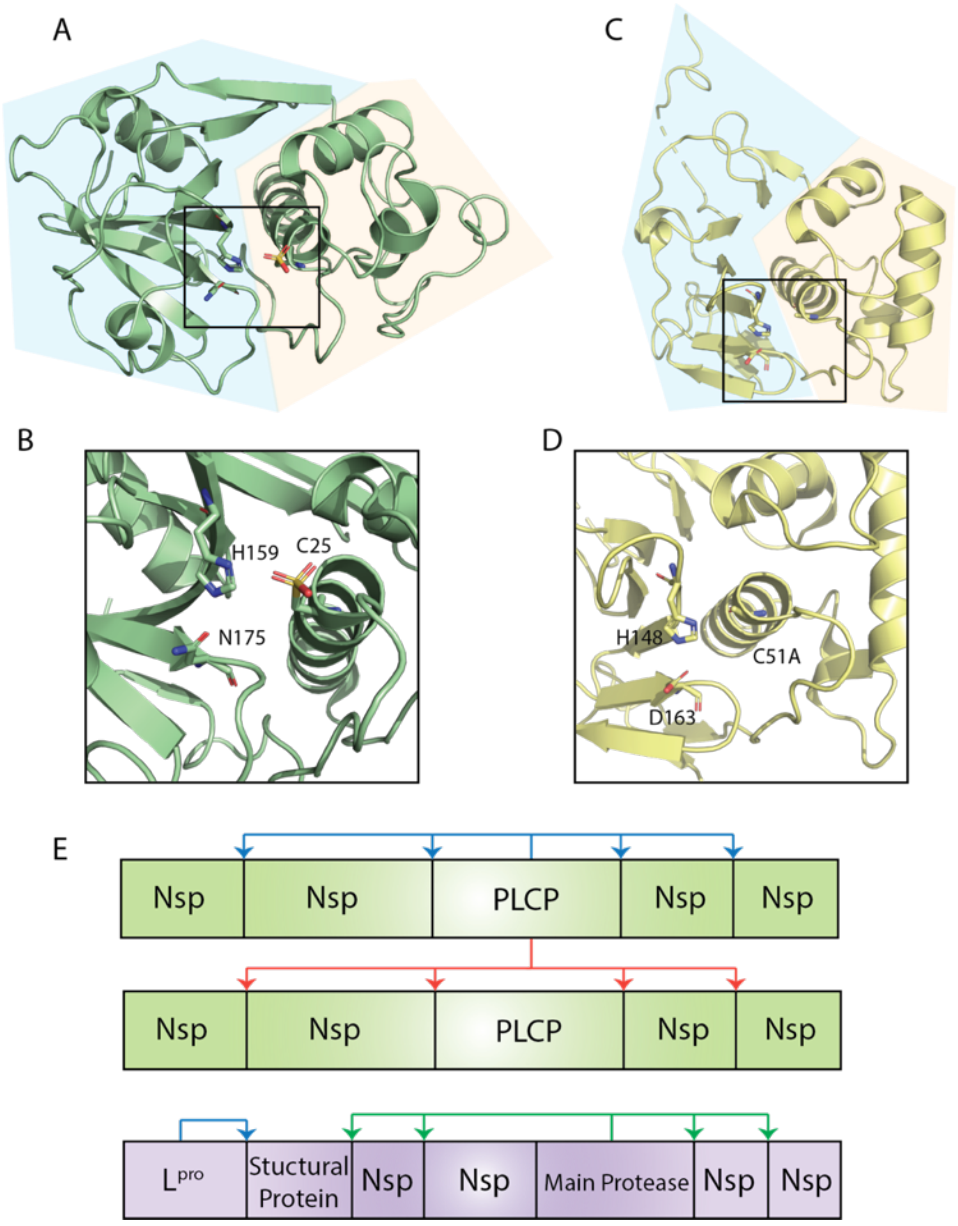


Figure 1.1 PLCP architecture and polyprotein processing.

A) Crystal structure of papain (PDB: 9PAP). L and R domains shown in blue and orange panelling, respectively. B) Closeup of papain active site with triad members labeled (Note that Cys is oxidized). C) Crystal structure of FMDV L^{pro} (PDB: 1QOL) with L and R domains shown identically to papain. D) Closeup of FMDV L^{pro} active site with triad members labeled. E) Examples of polyprotein processing events carried out by by PLCPs. Top, PLCP acting in *cis* on own polyprotein and also in *trans* on another (middle). Bottom, PCLP (e.g., L^{pro} or HC-PRO) located at N-term of a polyprotein only acting in *cis* on itself with a main protease facilitating other cleavages.

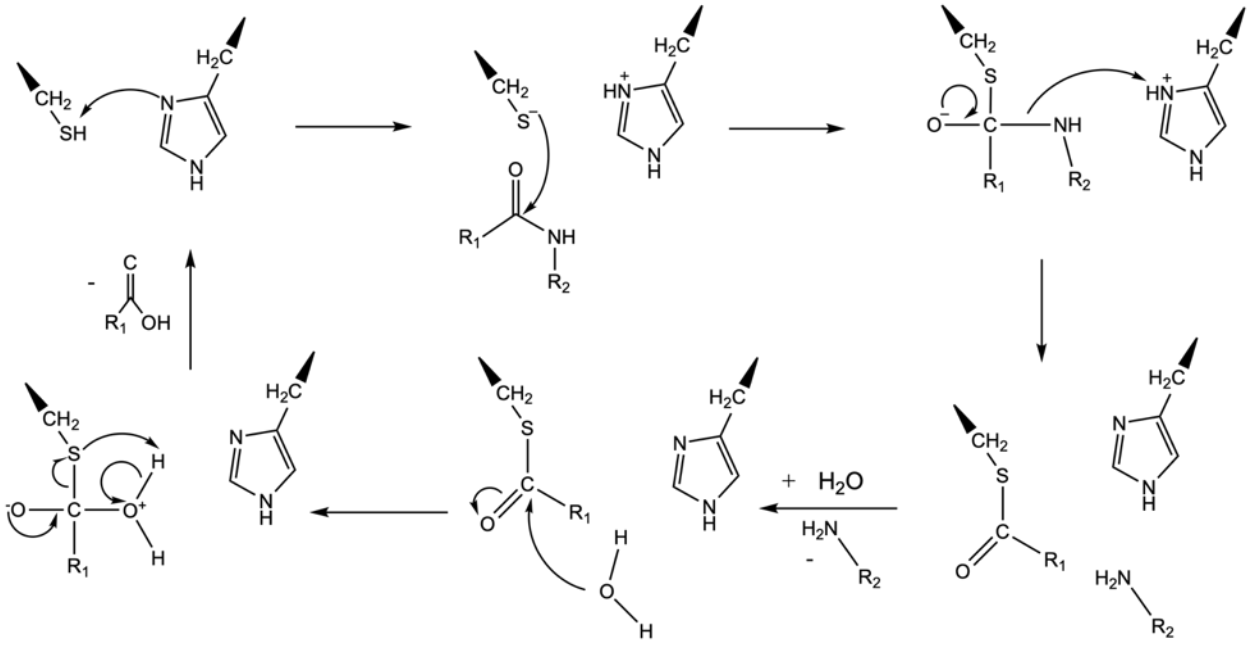


Figure 1.2 Classic cysteine protease mechanism featuring a catalytic dyad.

1.2.1 Traditional function of +ssRNA viral PLCPs

+ssRNA viruses depend on the synthesis of polyproteins, which are post-translationally cleaved into individual functional subunits. This mode of genome expression is common to these viruses, which aids in minimizing genome size by enabling the production of multiple proteins from a single open reading frame (ORF). This mode of protein production is also employed by these viruses since they undergo cytoplasmic replication and alternative RNA splicing strategies, processes typically occurring in the nucleus, are uncommon (10). The cleavage of polyproteins into their independent protein subunits is generally mediated by virus-encoded proteases (18, 19), however, cellular proteins have been shown to act as peptidases of +ssRNA viral polyproteins as well. This is the case with certain flaviviruses, such as Zika and dengue viruses in which endoplasmic reticulum- (ER) and Golgi-associated enzymes (signal peptidases and Furin, respectively) act to process the pr (premembrane), M (membrane) and E (envelope) structural proteins, which are essential processes for virion maturation (20–22). More conventionally, these viruses also encode for a viral chymotrypsin-like serine protease (non-structural protein 3 or NS3) that carries out the bulk of polyprotein processing (21).

Proteases from +ssRNA viruses come in many forms, such as chymotrypsin-like serine proteases, chymotrypsin-like cysteine proteases, papain-like cysteine proteases, ovarian tumor-like (OTU) cysteine proteases and pepsin-like aspartic proteases (19, 23–25). Quite often, multiple types of proteases can be found within the same viral polyprotein for its processing, or alternatively, the polyprotein is processed by one protease (18–20, 23, 26). These proteases can act in *cis* and/or *trans* to cleave their originator polyprotein as well as other polyproteins or act solely in *cis* on themselves sheerly for their own liberation (23, 27, 28).

1.2.1.1 Picornaviridae polyprotein and protease: The discovery of polyproteins and viral proteases

The first mention of the idea of viral polyproteins dates back to the 1960s in studies of viruses from the family *Picornaviridae* with the poliovirus and Mengovirus (29–31). In HeLa cells infected with type 1 poliovirus using a pulse-chase metabolic labelling experiment, ten total non-capsid proteins were detected from the cytoplasm of infected cells. It was known that the viral genome acts as mRNA (32) and that the genome size was ~6000 nucleotides long and thusly the protein to follow would be ~2000 residues in length. After short periods of radioactive amino acid incorporation, large proteins were detectable; however, over longer incubation periods, these larger proteins were almost undetectable and many smaller proteins were present (29). The investigators knew the virus contained only four capsid proteins (VP1-4), which should be detectable independently; however, ten noncapsid proteins were detectable from HeLa cell cytoplasm infected with poliovirus or Mengovirus (29, 30). Upon incorporation of phenylalanine, arginine, proline and methionine amino acid analogues in infected HeLa cells, a protein larger than ~200 kDa was the predominant species, linking some form of specificity regarding amino acid sequence and cleavage of the nascent polypeptide. Furthermore, when infected cells were treated with diisopropyl fluorophosphate (a known protease inhibitor), polypeptides of larger size were detectable in comparison to untreated cells, indicating polyprotein processing by a protease (33).

A later study was performed in reticulocyte lysate with encephalomyocarditis virus (a closely related picornavirus) in which viral mRNA was used in a cell-free system (34). Again, larger proteins were present during the early stages of translation; however, smaller proteins became apparent with time. To assess whether these proteins were an artifact of premature translational termination, edeine (an initiation inhibitor) was added (34). The studies were able to

see after the addition of edeine, smaller proteins were still detectable, indicating proteolysis was occurring by a factor specific to the presented mRNA. Additionally, upon the addition of N-ethylmaleimide (a cysteine protease inhibitor), smaller cleavage products were not detectable (34).

In 1981, studies were able to sequence the ~7.4 kb genome of poliovirus (35), which revealed that one single ORF spans ~90 % of the genome, further bolstering the idea of gene expression being centric around a polyprotein precursor and subsequent processing by internal viral proteases. When profiling the amino acid sequence, the dipeptide Gln-Gly was present 13 times, a disproportionately high occurrence assuming random distribution of residues and the size of the polyprotein. They theorized that this must be a signal sequence that is involved with protease recognition (35). Indeed, this dipeptide along with Gln-Ser, Glu-Ser, Glu-Gly, Ala-Asp, and Asn-Ser have been repeatedly implicated as cleavage motifs, showing a preference of viral proteases towards less bulky side chains (19). The main poliovirus protease was eventually described as being the 3C protease whose X-ray structure determined it to be a chymotrypsin-like Cys protease, although there is some debate about whether it also resembles papain (35–37). It was also found that poliovirus contains the additional chymotrypsin-like cysteine protease $2A^{pro}$, which cleaves the polyprotein between the first structural protein VP1 and $2A^{pro}$ itself; however, it has been shown that this cleavage ability is not essential to viral replication (38–40). The following sections will describe many different types of +ssRNA viruses from different viral classifications and specifically only those that contain PLCPs that act in polyprotein processing.

1.2.1.2 Alphavirus polyprotein and PLCP

Members of genus *Alphavirus* from family *Togaviridae* are vertebrate-affecting arboviruses (arthropod-borne) to infect humans, horses, birds and rodents (4). Notable members

of this genus include the human-affecting Sindbis, Semliki Forest and Chikungunya viruses (4). The alphavirus genome is split into two main ORFs, resulting in two major polyproteins. The first polyprotein encodes for four non-structural proteins (nsps) and the second predominantly hosts structural proteins (40, 41). The nsps expressed from ORF1 form the replicase machinery for the virus. Nsp2 is the only nsp to contain a proteolytic domain, which acts to process all three junction points between the nsps into their individual functional domains, while the structural protein-containing-polyprotein is processed by cellular enzymes (Furin and signal peptidases) in the ER and Golgi (43). As is the case with many +ssRNA viruses, RNA replication cannot occur until the RdRp (housed in nsp4) is released, showing the essential nature of this cleavage event (41). Similarly, cleavage of nsp1 (contains the methyltransferase and GTase domains for 5' capping) from nsp2 functionally results in a replicative shift from negative-sense RNA synthesis to positive-sense RNA synthesis (41). The cleavage order and nature of cleavage is finely regulated (17, 43, 44). Early in infection, cleavage carries out in *cis* between the nsp3|nsp4 junction of the first polyprotein. Cleavage events between nsp1|nsp2|nsp3 only occur in *trans* and thus occur later on during infection, meaning a polyprotein comprised of nsp1-nsp2-nsp3 and lone nsp4 dominate at early stages (an example of *cis*/*trans* cleavage is shown in Fig. 1.1E) (17, 43, 44). As the nsp1-nsp2-nsp3 polyprotein accumulates, cleavage by nsp2 from one polyprotein can act on another nsp1-nsp2-nsp3 molecule at nsp1|nsp2, followed by in *trans* cleavage at the nsp2|nsp3 junction to complete the cleavage cycle (17, 43, 44).

Work in the late 1980s was able to show that the C-term portion of nsp2 from the Sindbis virus contained a proteolytic domain (17). Experiments using blocking antibodies for each nsp only affected polyprotein processing when anti-nsp2 antibodies were used (17). Deletions in the N-term of nsp2 did not affect polyprotein processing, where C-term deletions greatly reduced

proteolysis, which could be restored by exogenously supplied viral proteins (17). Originally thought to be a metalloprotease, nsp2 was subsequently theorized to be a thiol protease, as reducing agents seem to be required for polyprotein processing during *in vitro* translation experiments (17). A later study on the Semliki Forest virus nsp2 domain confirmed this and also showed that proteolysis was inhibited by N-ethylmaleimide, but not by EDTA (metalloprotease inhibitor), pepstatin (aspartic protease inhibitor), or PMSF (serine protease inhibitor) (45). The functional studies on the protease domain of Sindbis virus nsp2 and sequence analysis and comparison with papain (11, 12) led to the first classification of a viral protease being thought of as papain-like (17).

1.2.1.3 Aphthovirus and Erbovirus polyprotein and PLCP

Aphthovirus is a genus of +ssRNA viruses within the family *Picornaviridae.* The most well-known and characterized member of the genus is the Foot-and-mouth disease virus (FMDV). FMDV was the first animal-affecting virus to be discovered in 1898 and infects common livestock, such as cattle, goats, swine and sheep (4). The most related genus is *Erbovirus*, which contains three horse-affecting viruses (4). There are two polyproteins produced from the ~8 kb +ssRNA genome of these viruses, one of which is produced via ribosomal skipping (4). These viruses encode for three proteases (46). The main 3C chymotrypsin-like Cys protease (termed 3C) is present in the second polyprotein produced from ribosomal skipping along with another chymotrypsin-like Cys protease ($2A^{pro}$), while the third protease, a PLCP, is present at the very N-term of the primary polyprotein and is termed the leader protease or L^{pro} and is considered an accessory protease (46–48).

L^{pro} is *cis*-acting (Fig. 1.1E) and only has one cleavage site at the junction of its own C-term and structural protein VP4, liberating it from the polyprotein. L^{pro} has been theorized as being

a PLCP for decades based on the location of the catalytic Cys-His as shown in Fig. 1.1C & D (49). L^{pro}, like papain, has a triad member in Asp163 (50). Members of the closely related genus *Cardiovirus* also have a version of L^{pro}, however, this protein does exhibit protease activity (46). Interestingly, L^{pro} can be found as two different variants (Lab^{pro} and Lb^{pro}) based on where translation initiation occurs, as two start codons are present 84 nt apart at the 5' end of the genome (46, 47, 51). Lab^{pro} is generated when translation commences from the most 5' AUG, however Lb^{pro}, where translation begins at the AUG 84 nt into the genome, is more abundant in infected cells (51). Both versions retain catalytic activity as PLCPs, however the loss of Lab^{pro} translation is much more deleterious than Lb^{pro}, the latter of which still results in viable virus capable of replication in animal models (51). The Lb form of L^{pro} from FMDV was the first viral PLCP to be structurally characterized, although as a catalytic knockout with a Cys51Ala mutation (Fig. 1.1C & D) (50, 52). The structure revealed similarity to papain, retaining the L and R domains and active site being formed at the interface of the two core domains (50).

Polyprotein processing is not the singular role of L^{pro}. Studies have shown that FMDV L^{pro} uses its protease activity towards cellular proteins as well, specifically eukaryotic initiation factor (eIF) 4GI (4, 46, 53, 54). eIF4F is the protein complex that recruits the ribosome to bind host mRNA for cap-dependent translation (55). The complex is made up of a number of proteins, including eIF4GI, which acts a scaffolding protein, using multiple domains to tether the complex together as well as being involved with 40S ribosomal subunit recruitment to the 5' end of capped mRNA (55). L^{pro} is able to cleave eIF4GI into two domain, disrupting eIF4F complex formation, inhibiting cellular cap-dependent mRNA translation, which does not affect FMDV and related viruses, as they are not 5' capped and use an internal ribosomal entry site for translation initiation, which in turn drives ribosomal activity predominantly towards viral protein production (50, 51,

53, 54). FMDV L^{pro} has also been shown to act as a deubiquitinase (DUB) to corrupt cellular signaling in the innate immune response to viral infection (51). Specifically, this ability results in the negative regulation and antagonism of the type I interferon (IFN) pathway (51). Similar activity will be discussed in detail in relation to nidoviruses.

1.2.1.4 Potyviridae polyprotein and PLCP

Members of the family *Potyvirirae* have plants as their natural hosts (4). They are the most biologically and agriculturally relevant family of RNA viruses to affect plants (56). The largest genus is *Potyvirus* with over 150 known species (4, 13). These viruses belong to the picornavirus-like supergroup based on their polyprotein arrangement and phylogenetic analysis of the RdRp sequences (23, 57). One staple of the picornavirus-like supergroup is the presence of a main 3C-like Cys protease, whose primary function is polyprotein processing; however, these viruses also have accessory leader proteases in the form of PLCPs found at the N-term of the polyprotein that are minorly involved with proteolytic processing and have other roles during infection (13, 23, 27). The main protease of potyviruses, NIa (nuclear inclusion protein a), has strict cleavage sequence requirements, only being able to cleave at Gln|Gly or Gln|Ser junction points (as above mentioned with poliovirus 3C) and typically has 6-8 cleavage sites in the primary polyprotein (13).

PLCPs have been theorized in a handful of potyvirids, specifically in the genera *Potyvirus* (Tobacco etch virus, Turnip mosaic virus, and Potato viruses A/Y), *Macluravirus* (Chinese yam necrotic mosaic virus) and *Bymovirus* (Barley yellow mosaic virus) (13). Although the locations of the coding sequence for these proteases in the genome are known, the function and classification for most are yet to be determined. However, based on the amino acid sequences of these accessory proteases, they have been thought of as being PLCPs for decades (49). The best characterized

PLCPs from the potyvirids are from the genera *Potyvirus* and *Rymovirus* and are termed the helper component protease (HC-Pro) (13, 23, 56). This protein is not shared amongst all potyvirids, and is exclusive to those with monopartite genomes, unlike members of genus *Bymovirus* who have bipartite genomes and have an alternative accessory protease termed P2-1, whose function is still relatively unclear (13).

HC-Pro is considered one of two minor proteases and solely cleaves itself in *cis* out of the lone potyvirid polyprotein at its own C-term, which is also done identically by the more N-terminally-located chymotrypsin-like serine P1-protease (58, 59). HC-Pro is comprised of three core domains, which are the N-term domain, central domain and the C-term domain, all of which have been shown to have various functions during potyvirus infection (27, 56, 58, 60, 61). The C-terminal region of HC-Pro contains the PLCP domain of the protein, composed of a catalytic Cys-His dyad and no triad member (27). Only one crystal structure of a potyviral HC-Pro PLCP domain exists and is from the Turnip mosaic virus (not to be confused with the Turnip yellow mosaic virus) (27). It has a minimalistic PLCP overall fold with its β-sheet rich domain being less pronounced; however, the two domains are present (27).

Three amino acid motifs of the N-term domain (DAG, PTK, and KITC) within HC-Pro have been shown to be involved directly with viral transmission (56, 58). Potyviruses are transmitted to their final plant hosts by aphids, typically as they are feeding on the inner leaf (56). The virus, which can replicate inside the aphid vector, is transmitted from the insect vector to the plant upon the physical puncturing of the aphid stylet into the leaf (56). Many studies have shown that the three N-term motifs of HC-Pro multimers facilitate the transmission of infectious virions to the plant host by assisting in viral particle association with the mouthparts of the aphid (56, 58). It has been suggested that these three amino acid motifs aid in forming a bridge between the viral

coat proteins of assembled viruses and receptors in the aphid stylet, acting to localize the viral particles to the mouthparts for heightened rates of transmission (56).

Another function of HC-Pro is that of tampering with the host antiviral silencing RNA mechanism by acting as an RNA silencing suppressor (RSS) (56, 58, 60). In plants, silencing RNA is one of the main modes of defence against RNA viruses. The process has been well studied and is triggered by the presence of dsRNA (replicating viral genome or secondary structures) and begins with the generation of small interfering RNA (siRNA) in the nucleus as double-stranded RNA, which is subsequently transported out of the nucleopore into the cytoplasm by exportin proteins (62, 63). The RNA is then recognized by Dicer-like protein(s) (DCLs) endoribonucleases, cleaving the dsRNA into shorter fragments typically 21-24 nt in length and then use their helicase domains to separate the dsRNA into ssRNA (64). Viral RNA can also directly act as template for siRNA (vsiRNA), which is directly accessible by host DCLs (64). The RNA-induced silencing complex (RISC) is then formed as the DCLs facilitate loading of the ssRNA onto Argonaute (AGO) proteins (65, 66). The RISC complex associates the ssRNA with the complementary viral RNA and AGO proteins carry out RNA degradation of the target (65–67). Host cell RdRps (present in plants) also may act to convert ssRNA to dsRNA, which can then be picked up by DCLs for subsequent RISC complex formation, amplifying silencing events (64).

HC-Pro from potyviruses and rymoviruses have been shown to act as RSSs by preventing loading of vsiRNA onto RISC (56). The 3’ ends of siRNA molecules are methylated for their protection by an RNA methyltransferase called HUA Enhancer 1 (HEN1), preventing degradation (68). Studies have shown that HC-Pro can inhibit methylation by preventing the production of the methyl group in the first place, which leads to siRNA degradation, or by obstructing the HEN1 methylation process altogether (64). HC-Pro can also act as an RSS by direct interaction with

AGO1, obstructing RISC formation, or by downregulating mRNA encoding for AGO1 by enhancing the production of an miRNA that targets AGO1 (miR168) (64). All of these roles combined make HC-Pro a truly multifunctional PLCP

1.2.1.5 Closterovirdae polyprotein and PLCP

Members from family *Closteroviridae* (type member: Beet yellow virus) have the largest genome size of plant-affecting +ssRNA viruses (4, 58). These viruses produce two polyproteins, with the one translated from ORF1a being substantially larger and also harbours the lone putative PLCP at the N-terminus termed L^{pro} (61, 69). As described with FMDV, closteroviruses have a leader protease located at the beginning of the major polyprotein. The catalytic domain is confined to the C-terminus region and is composed of a catalytic Cys-His dyad that only recognizes its C-terminus in *cis* at a Gly-Gly motif for autocatalytic release from the polyprotein (23, 70). Some closteroviruses encode for a second leader protease (L2) downstream of L^{pro}, however, its function is not well-known, but is believed to be involved in host range extension (58, 71).

This L^{pro} does not have DUB activity, but other accessory functions have been shown. As described with HC-Pro of the potyvirids, L^{pro} has been shown to have a role in genome amplification by acting as an RSS, presumably protecting the RNA genome from RISC-mediated degradation (61, 72). Another study showed by alanine scanning mutagenesis that the N-term region of L^{pro} is implicated in cell-to-cell movement of assembled viral particles within the phloem, implicating this protein's importance in long-distance transport (72).

As described above for many different viruses, PLCPs do not act solely as peptide cutters of the viral polyprotein. They often have very far-reaching functionality that acts to impart large advantages by modulating host cell response mediated by viral infection and also have purpose in

ensuring more successful and efficient replication and transmission. Taken together, it is no surprise that these enzymes are often targets for antiviral therapies. The auxiliary functions of other well-characterized PLCPs from members of the order *Nidovirales* and family *Tymoviridae* will be further discussed in the subsequent sections.

1.3 *Porcine reproductive and respiratory syndrome virus (PRRSV)*

1.3.1 Overview and implications of disease

PRRSV was first reported in the USA in multiple states in the late 1980s, as a large degree of reproductive failure was becoming increasingly evident in swine in the livestock industry and was first thought of as a mystery swine disease (73). These morbidities were also joined with respiratory abnormalities especially in newborns, stunted growth rates and increased mortality within the same herds (73–75). PRRSV also presents itself clinically as fever, lethargy, anorexia and physically as a blue discolouration of the ears (4). PRRSV was first successfully isolated by Dutch researchers from pregnant sows and piglets in 1991 where they were able to determine it was a unique etiological agent that was readily spread by aerosols (76). When PCR-based assays for detecting PRRSV became available in the 1990s, its presence was found virtually everywhere swine were kept for livestock purposes and highly virulent strains were being detected outside of North America and Europe (74, 77, 78).

The modes of transmission of PRRSV make spread particularly challenging to control. It had early been proposed that infection was through direct contact via uninfected and infected pigs, however, aerosols were soon after also shown to be risk factors (73, 75). Early findings described that PRRSV was detectable in virtually all porcine bodily secretions, making any form of secretion-based contact potential exposure routes (75). Transmission through semen is

predominantly problematic due to increased practices of artificial insemination, which increases the chances for long-distance transmission and is heightened by the extensive viability of infectious particles in semen (74, 79). Animal husbandry practices such as teeth clipping, tattooing, tail docking, and medicine administration are also possible modes of transmission through fomites (74, 79). Vertical transmission is highly problematic from pregnant sow to fetus and typically occurs in the last trimester of pregnancy (74).

PRRSV is one of the most economically important pathogens to infect swine globally. It has been estimated that PRRSV causes the US >$600 million per year in financial losses (80). It has been reported that the Canadian swine industry suffers ~$150 million per year due to PRRSV (81, 82). European estimates show that median losses in German swine farms moderately-affected by PRRSV lose ~€400,000 annually (80, 83). Unfortunately, literature-based evidence on treatment effectiveness has not been well-circulated, especially to farmers, making it challenging for farmers to justify and rationalize the cost of treatments over economic losses and return on investments (80).

1.3.2 PRRSV genome and polyprotein arrangement

PRRSV belongs to the family *Arteriviridae* in the order *Nidovirales* (4). There are two distinct species of PRRSV, PRRSV-1 and PRRSV-2, which have ~40 % genetic variation at the nucleic acid level (74, 84). PRRSV-1 is considered to dominate in Europe, while PRRSV-2 in North America and Asia (74). More recently, PRRSV-1 and -2 have been classified into their own genera as *Betaarterivirus suid 1* and *Betaarterivirus suid 2*, respectively (85). PRRSV, along with all arteriviruses, has a +ssRNA, monopartite genome that is ~15 kb in length and harbours 12 distinct ORFs as presented in Fig. 1.2A (86–89). Roughly three-quarters of the 5’ proximal genome

is composed of two ORFs (ORF1a and ORF1b) that encode for replicase proteins (88). The 3' region of the genome is composed of eight ORFs that are nested subgenomic mRNAs (sg mRNA) that give rise to the structural proteins, such as the glyco-, membrane, envelope and nucleocapsid proteins (87).

ORF1a is directly accessible by the host cell ribosome, generating the ~2500 amino acid polyprotein 1a (pp1a) (88). Pp1a is composed of 10 nsps, four of which have proteolytic ability to process pp1a (88, 90, 91). Nsp4 is the main protease, which is a 3C-like serine protease (His-Asp-Ser triad), acting as a classic chymotrypsin-like protease to cleave at six different junctions in pp1a and also acts to cleave the subdomains of proteins derived from ORF1b (91). Nsp1α (leader protease) and nsp1β are compact PLCPs with simplified active sites made up of catalytic Cys-His dyads, where nsp1α also contains a zing finger (ZF) domain (92–94). Nsp1α only acts on its own C-term and its liberation from pp1a is thought to be favoured in *cis*, as is the case with FMDV L^{pro} (93). It has been shown that nsp1α also has accessory functions in regulating sg mRNA synthesis for the production of the structural proteins (94, 95). Moreover, mutational analysis implicated the ZF domain in inhibiting the production of IFN-β, a crucial signaling cytokine involved in antiviral defence mechanisms of the immune system (96). Nsp1β is structurally similar to nsp1α, but lacks an N-term ZF domain (92). The protease is also responsible for only one cleavage at its C-term and mutations in the active site result in lack of genomic replication (97). Nsp1β and its auxiliary roles will be explored in detail in sections to follow.

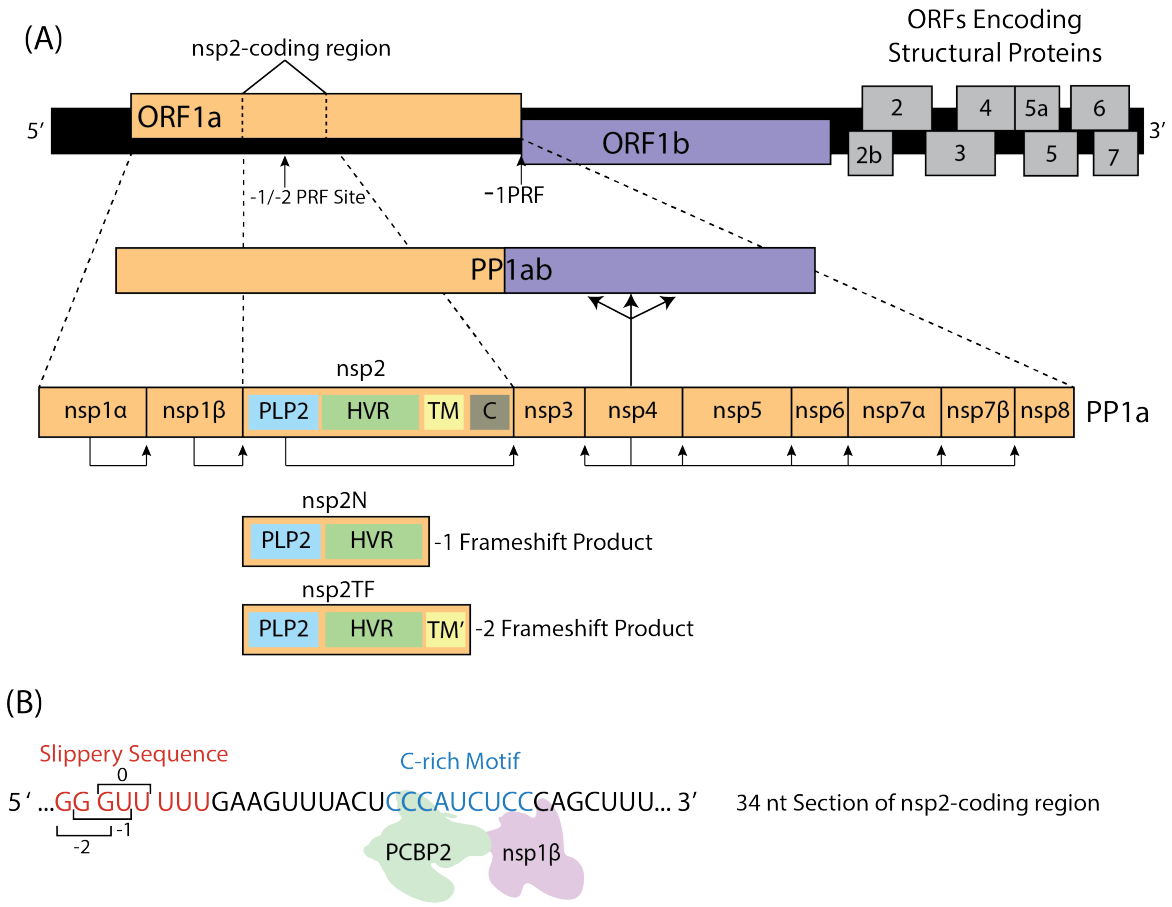


Figure 1.3 PRRSV genome, polyprotein and PRF events.
A) schematic of the full +ssRNA PRRSV genome. 3' ORFs encoding structural proteins are shown in grey. Translation of the largest ORFs (ORF1a and ORF1b) yields replicase polyproteins 1a and 1ab (pp1a and pp1ab). Translation of ORF1b requires a -1 PRF event at the end of ORF1a. PRRSV pp1a is comprised of 10 nonstructural protein (nsp) subunits, four of which have autocatalytic polyprotein cleavage activity (arrows indicate cleavage sites). Whereas the –1 PRF at the ORF1a/1b junction is directed by stimulatory RNA structures, additional -1 and -2 PRF events occurring within the nsp2- coding region do not depend on higher-order RNA structures, additional -1 and -2 PRF events occurring within the nsp2- coding region do not depend on higher-order RNA structures. These PRF events result in truncated nsp2 variants, nsp2N and nsp2TF, which both retain the papain-like cysteine protease (PLP2) and hypervariable regions (HVR) but lack the C-terminal Cys-rich domain (C). Nsp2TF also contains a modified transmembrane domain (TM') that is encoded by a short alternative ORF (TF) that overlaps with ORF1a in the -2 reading frame. B) The region of ORF1a from PRRSV SD01-08 showing where -1/-2 PRF occurs within nsp2. The slippery sequence is shown in red, and the C-rich motif is shown in blue. The C-rich motif replaces the canonical higher-order RNA structural element as found in most other PRF mechanisms and serves as a putative binding site for nsp1β and PCBP2. Adapted from (98, 99).

Nsp2 is by far the largest protein produced by PRRSV at roughly 1100 residues in length. The N-term portion contains a PLCP domain termed PLP2 (87, 88). The other domains within nsp2 are the hypervariable region (HVR), transmembrane domain (TM), and Cys-rich domain (28, 88). The nsp2-coding region of pp1a has the highest proportion of genetic variability amongst PRRSV isolates, with deletions resulting in a loss of 150 amino acids being reported in certain PRRSV isolates (100). The crystal structure of PRRSV PLP2 is yet to be reported, however, the corresponding domain of the related Equine arteritis virus (EAV) has been determined (101). The protease has a catalytic triad (Cys-His-Asn) similar to papain and also contains a ZF domain (101). Closer structural comparison revealed PLP2 to be more structurally analogous to the OTU superfamily (101). There is only one cleavage event carried out by the enzyme at the nsp2|nsp3 junction, which has been shown to be essential to viral replication (102).

PLP2 demonstrates significant capacity to modulate the innate immune response of the infected host by having DUB activity and the ability to hydrolyze interferon-stimulated gene 15- (ISG15-) conjugated proteins, acting to corrupt cytokine signaling pathways (103). Specifically, the ZF domain is crucial in recognizing ubiquitin (Ub) and ISG15 (101). Interestingly, nsp2 exists as multiple variants or isoforms in infected cells, which will be discussed in detail in sections to come (89, 104, 105); however, one of these isoforms has also been connected in association with the nucleocapsid and may also act as a structural protein with unknown function(s) (84, 104).

1.3.3 The various roles of PRRSV nsp1β

PRRSV nsp1β was first characterized from a European strain of PRRSV-1 (106). The homologous protein in EAV was shown to autocatalytically cleave at the nsp1|nsp2 junction and

sequence analysis classified the domain as a PLCP (107). Indeed, similar domains were found in PRRSV and the lactate dehydrogenase-elevating virus (LDV) with conserved locations of putative active site Cys-His residues (106). *In vitro* translation assays using rabbit reticulocyte lysate confirmed that both PRRSV nsp1α and β have proteolytic processing ability towards their C-term (106). Mutational experiments centric towards nsp1α/β active site Cys or His significantly reduced or abolished protease activity, and as previously shown with EAV, both proteases seem to only act in *cis* (106, 107).

Using a PRRSV reverse genetics system, the importance of nsp1α and β in the context of the PRRSV life cycle was explored by the same group (108). Active site mutations in nsp1α allowed for efficient genome replication but did affect sg mRNA synthesis, whereas analogous mutations in nsp1β prevented genome replication, as no detectable RNA synthesis occurred in infected cells, suggesting that it has more of a function than just acting to cleave at the nsp1β|nsp2 junction (108). As described above for the potyvirus HC-Pro and FMDV L^{pro}, proteases at the N-term of the polyprotein that typically act to only cleave themselves from the polyprotein often have more global and important roles in the context of viral replication and infection.

More recently a new role for nsp1β was described, involving the "imprisonment" of cellular mRNA in the nucleus of MARC-145 cells infected with PRRSV-1 and -2 strains (109). It is believed that this would stifle mRNA transport to the cytoplasm for translation, allowing PRRSV to exploit translational machinery for itself and also potentially acting to supress expression of host innate defense proteins (109). Immunofluorescence assays coupled with quantitative reverse transcription-PCR (qRT-PCR) for the viral N protein transcripts as well as four cellular gene transcripts were carried out, which showed that in PRRSV infected cells, the ratio of the cellular transcripts were significantly higher in the nucleus in comparison to the cytosol relative to

uninfected cells, whereas, as expected, N protein transcripts were mostly cytosolic (109). When individual viral gene products (nsp1α/β or N protein) were delivered to cells, only cells transfected with plasmid coding for nsp1β elicited evident mRNA retention to the nucleus (109). More specifically, an nsp1β SAP motif (**S**AF-A/B, **A**cinus, **P**IAS; SAF-scaffold attachment factor, PIAS-protein inactivator of activated STAT, STAT-signal transducer and activator of transcription) spanning residues ~123-138, which is well-conserved in both types of PRRSV, has been implicated in this mRNA "imprisonment" (109). SAP motifs are often found in nuclear proteins involved with DNA repair, RNA processing, transcription, and other roles (110) and have been found in certain RNA viral proteases such as FMDV L^{pro} (111).

An alternative function of nsp1β has been well-explored, relating to transactivating programmed ribosomal frameshifting to open two new ORFs in the nsp2-coding region of ORF1a, resulting in two variants or isoforms of nsp2, termed nsp2TF and nsp2N (Fig. 1.2A) (89, 105). Specifics on how this PRF event occurs will be discussed in detail.

1.3.3.1 Canonical programmed ribosomal frameshifting in viruses

Viruses heavily depend on their host cell machineries to express all necessary proteins and form viral particles. As explored above, the production of polyproteins is a way for the virus to pack substantial genetic information into single ORFs for the generation of multiple proteins. However, +ssRNA viruses may also utilize other non-canonical translation means to produce their proteins, which is often guided by the structure and sequence of their genomes (10). These mechanisms of translation are clever ways for viruses to regulate protein expression levels, as well as increasing the coding capacity of the genome (10). Being able to produce multiple proteins from a single genome via polyprotein production and non-canonical translation is advantageous to

viruses with non-segmented genomes, as packing of the genome into viral particles becomes simplified and less dependent on ensuring all segments have been delivered appropriately (10).

One non-canonical translation method often used by +ssRNA viruses is programmed ribosomal frameshifting (PRF). PRF was first described for the Rous sarcoma virus where overlapping ORFs encoding for the Gag and Pol proteins undergo a PRF event to produce a Gag-Pol polyprotein (112, 113). A stop codon exists at the 3' of the *gag* ORF, which typically prevents further translation into the *pol* region; however, translation into *pol* does occur occasionally to produce the fusion protein at a ratio of 20:1 (Gag to Gag-Pol), which is facilitated by a -1 PRF event (112). The presence of a heptameric, "slippery" frameshift site with sequence X_XXY_YYZ (X any three identical nucleotides, Y either repeating A/U, and Z is A/C/U) has been described for Rous sarcoma virus and related viruses and has been shown to be one essential element in stimulating PRF; however, not alone by itself (10, 112). An immediate downstream RNA secondary structure in the form of a stem-loop (or pseudoknot) was found to be another vital motif in eliciting PRF (112, 114). These two stimulatory RNA motifs are essential in viral canonical PRF, as is the case with pp1ab translation in PRRSV (10, 115).

It is believed that upon translation of the viral genome by the host cell ribosome, the ribosome stalls when it encounters the RNA secondary structure, which acts as a kinetic obstacle as the ribosome must unwind the element to continue translation (10, 114). The ribosome is typically located on the upstream slippery sequence as it is unfolding the RNA, and the nature of this sequence acts to facilitate PRF (typically in the -1 frame) (10). It has been hypothesized that the ribosome is associated with peptidyl and aminoacyl tRNAs at this point and detachment from the zero-frame codons and subsequent reattachment in the -1 frame (presuming PRF occurs) still allows for near-perfect realignment (outside of the wobble position) of the new frame due to the

nature of the slippery sequence (i.e. X_XXY_YYZ (0 frame) → XXX_YYY_Z (-1 frame)) (10, 112). Should the frameshift occur, the ribosome will continue to translate in the new ORF until termination through stop codon detection, allowing for access to multiple overlapping ORFs within one stretch of RNA (10, 116). The ribosome does not always slip during each pass and often remains in the zero-frame during unwinding, however, PRF events do occur at various efficiencies depending on the stimulatory elements in question (10, 117, 118).

As seen in all nidoviruses, there is a natural stop codon in the 0-frame in between ORF1a and ORF1b that prevents readthrough and subsequent translation into ORF1b, however a -1 ribosomal frameshift at this junction silences the stop codon and translation can occur into ORF1b to produce the large polyprotein pp1ab (119). This frameshift occurs roughly 15-20 % of all ribosomal passes (using a reporter system) and allows for the formation of nsps9-12 (87, 120). This efficiency is significantly lower than related +ssRNA viruses like the encephalomyocarditis virus of family *Picornaviridae* where a -1 PRF event occurs at ~70 % efficiency, albeit having relatively different genomic architecture (121). As above mentioned, the subdomains produced from translation into ORF1b are processed by nsp4 to give more key replicase proteins, such as the RdRp and the RNA helicase, however, these proteins from PRRSV will not be discussed herein. Although PRRSV uses canonical PRF at the ORF1a|ORF1b junction, more recently, it has been found that a novel form of -1 and -2 PRF also occurs in the nsp2-coding region of ORF1a, which is transactivated by nsp1β and cellular proteins, Poly(C)-binding proteins 1 and 2 (PCBP1/2) (105, 122, 123).

1.3.3.2 Alternative PRF events in PRRSV

A previous study analyzed the nucleotide sequence of ORF1a for >200 isolates of PRRSV in 2012 and observed some peculiarities in the nsp2-coding region in all isolates that suggested a -2 PRF event in this region (105). Of all the regions of PRRSV ORF1a, this region of the genome clearly has the highest overall conservation with the most reduced rate of synonymous substitutions, suggesting the importance of the nucleic acid sequence for this stretch of the genome (105). When sequences for all isolates were shifted *in silico* in the + 1 frame (or -2 frame), they noticed an absence of stop codons for ~170 codons in this region, which was also shown to be shared with similar viruses LDV and Simian hemorrhagic fever virus (SHFV), but not in EAV (105). A -2 frameshift would produce an nsp2 variant protein (nsp2TF) which shares identical PLP2 and HVR domains, but the -2 PRF occurs just before the transmembrane domain, creating a variation stemming from this overlapping ORF with an alternative TM domain termed TM' (105). A stop codon in the -2-frame is present just downstream the TM'-coding region, resulting in a truncated version of nsp2 with an alternate membrane-spanning region and no Cys-rich region (Fig 1.2A) (105). Located at the 5' end of the newly found ORF in almost all PRRSV isolates analyzed (206 of 212), the LDV sequences, and one SHFV sequence, a G_GUU_UUU motif is present, having all criteria of a typical heptameric slippery sequence (Fig. 1.2B) (105). Interestingly, no downstream RNA secondary structure could be predicted, although a highly conserved cytosine-rich motif 11 nt downstream of the slippery sequence was present (Fig. 1.2B) for PRRSV, SHFV and LDV isolates analyzed (105).

Immunodetection using antibodies specific to the N-terminus of nsp2 or the TM' portion of nsp2TF in MARC-145 cells infected with PRRSV-1 or PRRSV-2 strains confirmed the presence of nsp2TF as well as an even more truncated version of nsp2/nsp2TF termed nsp2N. Residues

comprising the TM' domain of nsp2TF, were confirmed using mass spectrometry (MS) (105). Nsp2N would arise from a -1 PRF event at this same site; however, there is an immediate stop codon following the slippery sequence in this frame, which would lead to a protein with only the PLP2 and HVR domains and no membrane-spanning domain (Fig. 1.2A). Transient expression of ORF1a using a recombinant vaccinia virus/T7 polymerase expression system confirmed that mutations in the shift site (G_GU**A**_UU**C**) or in the C-rich motif significantly hindered frameshifting, reducing nsp2TF and nsp2N levels, while nsp2 levels went up considerably, as ribosomes are presumably not being directed into these alternate frames (105). PRF efficiencies were estimated to be ~15 % and ~6 % for -2 and -1, respectively in a wild type (WT) scenario (105).

1.3.3.3 Nsp1β transactivates -1/-2 PRF in the nsp2-coding region

As previously mentioned, delivering PRRSV ORF1a to cells is sufficient to stimulate -1/-2 PRF, so authors in a subsequent paper wanted to define the minimal sequence of ORF1a to produce PRF products (89). Using the same expression system as mentioned above, a study found that delivery of constructs absent of nsp1α or nsp3 still promoted -1/-2 PRF product formation, whereas constructs lacking nsp1β expressed nsp2, but significantly less amounts of nsp2TF/nsp2N (89). Further, nsp1β introduced on the same construct with nsp2 in comparison to on an independent construct also allowed for production of PRF products, suggesting PRF-mediated by nsp1β also occurs in *trans* (89).

When essentially every codon of nsp1β was mutated synonymously, PRF still occurred, indicating that the protein and not the RNA coding for nsp1β is implicated in PRF (89). An RNA-binding helix in nsp1β with sequence GKYLQRRLQ is highly conserved in both types of PRRSV

and contains many basic residues that are often involved with nucleic acid binding (89, 92). Mutational analysis in this RBM showed the that this helix is essential in stimulating PRF (89). In cells co-transfected with constructs coding for nsp1β and a plasmid that would result in the production of RNA that represents 79 nt of the PRRSV genome (containing the slippery sequence and C-rich motif), nsp1β was pulled down from cellular lysate using an α-nsp1β Ab, followed by qRT-PCR of the RNA that co-immunoprecipitated (co-IP) with nsp1β (89). These findings suggested that nsp1β has a propensity to bind the PRRSV genome at or close to the shift site (89).

Previous studies on PRRSV showed that nsp1β assays involving co-IP followed by MS resulted in the pulldown of cellular proteins PCBP1 and PCBP2 in cells transfected with a construct coding for nsp1β (124). In MARC-145 cells infected with PRRSV, immunofluorescence determined colocalization of nsp1β and PCBP1/2 (124). In the same study, siRNA targeted towards PCBP1, PCBP2 or a combination of both significantly reduced PRRSV titers in MARC-145 cells 24 h post-transfection, with lowest titers founds for PCBP2 siRNA in comparison to PCBP1 and the lowest in the combination treatment (124). These findings were confirmed by another group showing similar nsp1β/PCBP2 interactions using a yeast-two hybrid system and immunofluorescence studies in MARC-145 and PAM (porcine alveolar macrophage) cells transfected with a construct coding for nsp1β or infected with a strain of highly-pathogenic (HP)-PRRSV, a strain more similar to PRRSV-2 (125).

1.3.3.4 Roles of nsp1β and PCBP1/2 in PRRSV PRF

PCBPs have been shown to be involved with mRNA stabilization as well as transcriptional and translational enhancement (126, 127). To date, there have been four PCBPs characterized (PCBP1 to 4), all of which have a high-tendency of binding cytosine-rich nucleic acid (126).

PCBP1 and 2 have been the best characterized, specifically their K-homology (KH) domains. Both PCBP1 and 2 have three KH domains, all of which have been implicated in nucleic acid binding (128, 129). Studies in 2016 revealed that PCBP1 and 2 stimulate PRRSV PRF in conjunction with nsp1β to form nsp2TF and nsp2N, showing for the first time a cellular protein transactivating viral PRF (123). Previous findings were able to show the importance of the C-rich motif of the PRRSV genome in frameshifting and also the significance of the linker length between the slippery sequence and the C-rich motif in stimulating PRF in reticulocyte lysate (123).

Electrophoretic mobility shift assays (EMSAs) using recombinantly produced nsp1β, PCBP1, and PCBP2 showed that there was synergistic binding of nsp1β/PCBP1 or nsp1β/PCBP2 to a 58 nt RNA probe representing the genome, containing the slippery sequence and C-rich motif (123). Mutational analysis within the KH domains was carried out at the important, shared GXXG loop (to GDDG), which has been shown to affect RNA binding of the KH domains without affecting stability and folding, as determined by circular dichroism (123, 130). These studies showed that PCBP1/2 GDDG mutations in KH1 and KH3 significantly impacted PRF product formation and complex formation through EMSAs, but no change was seen between WT and GDDG mutations with KH2, suggesting KH2 of either PCBP1/2 is not important in stimulating PRF (123).

1.3.3.5 Roles of nsp2TF and nsp2N in PRRSV infection

Nsp1β and PCBP1/2 have convincingly been shown to stimulate PRRSV PRF, but the implications of producing the nsp2 variant proteins was not well understood until 2018 (131). As mentioned above, the PLP2 domain of nsp2 has been shown to have significant impact on supressing the host innate immune response by inhibiting the type I IFN pathway by disrupting

cellular protein signaling using the DUB and de-ISGylating activities of PLP2 (101, 103, 132). Interestingly, both nsp2TF and nsp2N still contain the PLP2 domain, suggesting they too may be involved in similar functions.

Studies were able to show *in vitro* that knocking out the active site Cys/His for nsp2TF and nsp2N significantly reduced the global ability for either protein to decouple Ub or ISG15 from cellular proteins, demonstrating that these two proteins also harbour DUB and de-ISGylating activities (131). Immune gene mRNA profiling of MARC-145 and PAM cells infected with WT PRRSV, PRRSV that produces nsp2TF without TM' (terms KO1), or PRRSV that can only produce nsp2 (termed KO2) determined that there was a significant upregulation of immune genes in PRRSV KO1- and KO2-infected cells, specifically in those involved with cytokine receptor interaction and tumor necrosis factor signalling (131). These results were confirmed using *in vivo* studies in naïve pigs infected with all three strains of PRRSV and by analyzing serum from the specimen six days post-infection (131). Reduced viral load was seen in pigs infected with PRRSV KO1/KO2 compared to WT PRRSV and mRNA levels of interferon genes and counts of T cells in PRRSV KO1/KO2-infected pigs were significantly higher, suggesting the importance of nsp2TF and nsp2N in corrupting the innate and adaptive immune responses (131).

1.4 *Viruses of the family Tymoviridae*

1.4.1 Overview: hosts, infection and disease

Viruses from the family *Tymoviridae* are plant-affecting viruses subdivided into three genera: *Tymovirus, Marafivirus* and *Maculavirus*, which have been grouped together based on the sequence similarity of their RdRps and coat proteins (CP) (133). Tymoviruses and marafiviruses have been far better studied over the years and will be the focus of discussion. There are currently

30 species within the genus *Tymovirus*, all of which infect exclusively dicotyledonous plants (133, 134). There are currently only 10 species members within genus *Marafivirus*, with new purported members still being reported, yet marafiviruses can infect both dicots and monocots, with many hosts being grasses/cereals (133–136). The type members for both genera are the Turnip yellow mosaic virus (TYMV) and Maize rayado fino virus (MRFV), *Tymovirus* and *Marafivirus*, respectively (133). The amount of research done on plant-affecting viruses is noticeably far less than on animal- and human-affecting viruses, so it follows that details on general replication strategies and transmission is not abundant, however, there is still enough scientific literature to summarize these topics, with substantial amounts of work done on the molecular biology of TYMV over the years.

Tymovirids are naturally transmitted to host plant tissue (typically leaves) by insect vectors, although can be readily transmitted through vascular puncture of seed/leaf tissue or via leaf rubbing in a lab setting (133, 137–139). Tymoviruses are generally spread by beetles, whereas marafiviruses by leafhoppers (133, 137). Tymoviruses are thought to be semi-persistent, non-propagative viruses in that they are not internalized deep into the insect vector system and likely remain in the foregut and mouthparts (133, 140). Interestingly, marafiviruses are persistent-propagative viruses in that they are internalized far into the midgut of the insect vector and are able to replicate in their vector as well (133, 137, 140). Of note, marafiviruses are only limited to phloem cells unlike tymoviruses which can spread to multiple tissues (141).

Viruses from *Tymoviridae* have a virtually global distribution (142). These viruses affect overall plant health quite dramatically in that they lead to significant defoliation, visible chlorosis that results in inefficient photosynthesis, stunted growth, necrosis of sieve tubes and phloem cells and reduction in fruit bearing (135, 143–145). These infections can be extremely problematic in

the field if bacterial or fungal pathogens are also present, compounding impacts on the crops (146). These are problematic viruses as they affect many crops essential in food and alcohol production as well as those relevant to biofuels. The heightened spread of these viruses, increasing global population leading to increased food demand, coupled with climate control issues that are affecting agriculture practices globally (147) underpin the importance of better understanding these viruses from a molecular standpoint, which will be discussed further (148–150).

1.4.2 Tymoviridae genome and proteome

The genomes for TYMV and MRFV were sequenced in the late 1980s and early 2000s, respectively (146, 151). They have remarkably similar genome sizes at ~6.3 kb that only differ by 13 nt in the two isolates that were first sequenced (146, 151). The genomes are both 5' capped and are arranged very similarly with one predominant ORF that spans roughly 90-95 % of the genome as depicted in Fig 1.3 (146, 151). This major ORF is translated directly by the host cell ribosome to generate the replicase complex that is made up of four subdomains: the methyltransferase domain (MET), PLCP domain (PRO), helicase domain (HEL) and RdRp (141).

Both MRFV and TYMV have movement proteins termed p43 and p69, respectively, which give the viruses the ability to spread from cell to cell and throughout the tissues of the infected host (152). Marafiviral genomes actually code for two distinct coat proteins, a major and minor version (CP and CP*, respectively). CP is translated from the primary ORF that produces the replicase, resulting in a large polyprotein that contains the replicase fused to CP at the C-term of the RdRp, whereas CP* is produced from its own ORF via sgRNA in which a core sequence known as the marafibox serves as its own transcriptional promoter (152–154). This strategy for structural protein production differs from the tymoviruses, where only one coat protein is produced similarly

to CP* via sgRNA and a sequentially similar promoter termed the tymobox (155). Another difference between marafiviral and tymoviral genomes is the presence of a tRNA-like secondary structure at the 3' end of the tymovirus genome in place of a poly-A tail, as seen in the marafivirus genome (141, 156). This tRNA-like structure was first described for TYMV but has been since characterized for many other plant-affecting viruses and structurally resembles eukaryotic tRNA and is believed to be involved with binding eukaryotic elongation factor 1A, which normally is involved with delivering aminoacyl tRNAs to the ribosome, likely enhancing translation efficiency for TYMV (156, 157).

Although not the case for marafivirus PRO, the PRO domain of TYMV has been very well-characterized over the years and has been shown to have various functions including polyprotein processing, facilitating viral protein localization in the infected host, and DUB activity for modulating the host cell innate immune response (148–150), which will be discussed in the next section, however, one of the primary purposes of this thesis is to shed light on the marafiviral PRO domains, which will be discussed in future sections.

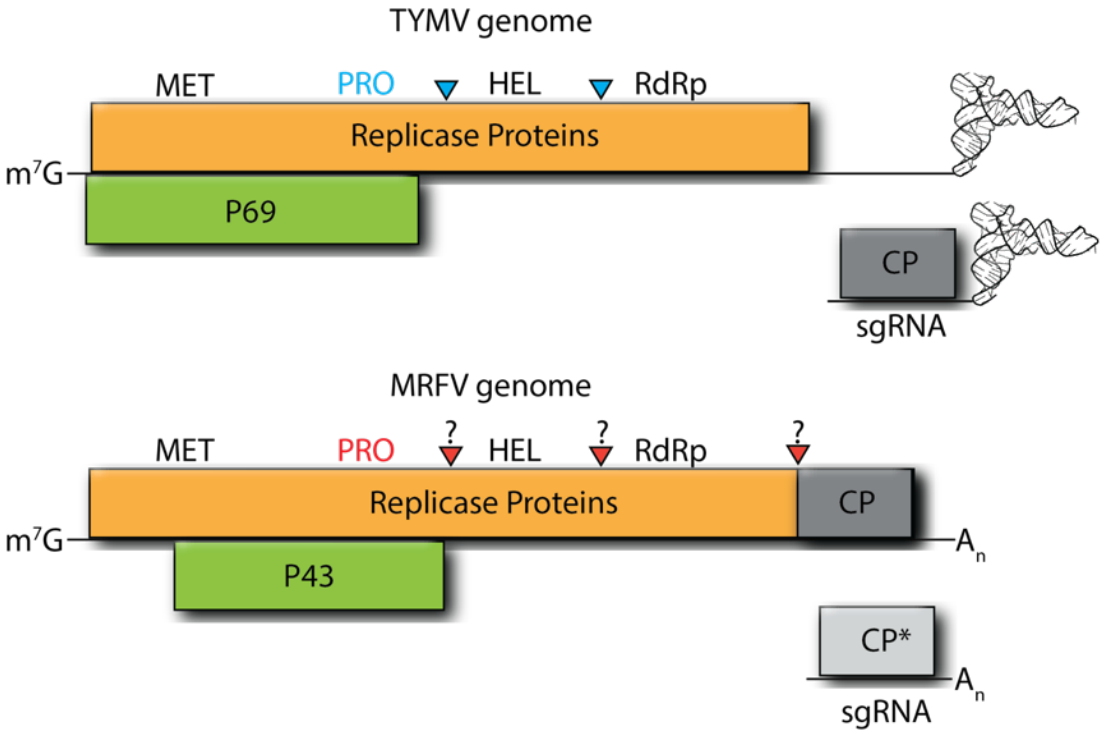


Figure 1.4 Tymoviridae genome arrangement and proteins encoded therein.

Tymovirus TYMV genome (top) and marafivirus MRFV genome (bottom). The Major ORFs are shown in orange, which encodes for the replicase proteins: methyltransferase (MET), protease (PRO), helicase (HEL), RNA-dependent RNA polymerase (RdRp), as well as the major coat protein for MRFV (CP). CP* is the minor coat protein produced only in MRFV through sgRNA, which is also how the sole CP for TYMV is produced. Movement proteins are shown in green ORFs, P69 and P43 for TYMV and MRFV, respectively. Known cleavage events by TYMV PRO shown with blue triangles and hypothesized cleavage events shown with red triangles by MRFV PRO. TYMV tRNA-like structure from PDB ID: 4P5J. Figure adapted from (152).

1.4.3 Roles of TYMV PRO

Preliminary studies began by investigating the gene expression of TYMV in the 1980s, shortly after the sequencing of its genome. The first thing that was apparent was the lack of a stop codon in a huge stretch of the genome (major ORF), which should be translated to give a protein of ~206 kDa; however, in an *in vitro* translation system using genomic RNA extracted from viral particles, no protein representing the full translation of the major ORF was detectable, yet smaller molecular weight proteins were, which led to the belief that some form of polyprotein processing was occurring (158). A few years later, another group using the same system showed that delivery of various cysteine protease inhibitors greatly reduced the formation of smaller molecular weight proteins, which was not seen for serine protease inhibitors (159). A subsequent paper using an infectious TYMV cDNA clone of the genomic RNA, an *in vitro* translation system, and antibodies specific to the termini of the major polyprotein, which results in two distinct protein species of ~70 kDa (RdRp) and ~150 kDa (MET-PRO-HEL) being generated, presumably from the 206 kDa precursor (148). This group was also able to determine the relative location of the protease domain and the cleavage site between the polyprotein|RdRp junction and mutational analysis showed that mutations altering the site greatly impacted cleavage efficiency (148). These findings were validated by the same group using turnip protoplasts (160). This was further confirmed by another paper that showed the importance of the catalytic Cys and His for cleavage to occur and also characterized the protease as a PLCP based on sequence comparison (161, 162).

In more recent years, tremendous work on TYMV PRO has been done by the Jupin group. They were able to generate antisera specific to the C-term of the helicase domain and a proline-rich region in between MET-PRO (163). Using their antibodies and a transient gene expression system in *Arabidopsis thaliana* protoplasts, they were able to determine that an additional

polyprotein cleavage event occurs at the PRO|HEL junction at the proteases own C-term to generate MET-PRO and HEL (163). Studies showed that this cleavage can also occur in *trans* in plant cells when the full-length major ORF with a catalytic Cys to Ser mutation was co-transfected with a construct only coding for WT PRO in *A. thaliana* protoplasts (163). Mutational analysis at the PRO|HEL junction confirmed the cleavage site to be in between the two domains (163). Northern blotting specific to minus-strand viral RNA confirmed that altering the cleavage site results in significantly decreased levels of detectable RNA, suggesting the importance of this cleavage in viral infection, specifically genome replication (163).

Previous studies focused on the organization of the replication complex and especially the RdRp and its targeting to chloroplasts (150, 164). Preliminary findings suggested that some portion of the MET-PRO-HEL polyprotein facilitated targeting of the RdRp to *A. thaliana* protoplast chloroplast envelopes and that there is no means of localization by the RdRp alone (165). *In planta* studies using Chinese cabbage infected with TYMV and fractionation of membrane structures revealed the presence of the RdRp and the MET-PRO-HEL polyprotein following western blotting specifically in the envelope of chloroplasts, as no significant signal was seen in the stroma or thylakoids (164). These targeting events were subsequently attributed to the PRO domain and has been suggested to be important in enhancing genomic RNA replication by physically trapping the RdRp as well as protecting viral RNAs from cellular siRNA machinery (164, 166).

The next area of research that emerged involved the DUB activity of TYMV PRO. In a 2010 publication, using a cell line of transgenic *A. thaliana* that expressed the RdRp, the RdRp was readily degraded in cells over time, which was slowed down by the addition of MG132 and clastolactacystin β-lactone, two known proteasome inhibitors (167). In a reporter based assay using a transient gene expression construct producing the RdRp fused to chloramphenicol

acetyltransferase (CAT), Ub and luciferase (LUC) to produce a CAT-Ub-RdRp-LUC fusion protein, luciferase levels were decreased significantly and rescued with the addition of the proteasome inhibitors, suggesting degradation by the 26S proteasome, a process that is controlled by Ub (167, 168). It was subsequently determined that the RdRp alone was ubiquitinated in *A. thaliana* naturally and that several lysine residues (the attachment residue of target proteins for Ubiquitin) of the RdRp are Ub-acceptors (167). It was also found that that co-expression of MET-HEL-PRO stabilized degradation of the RdRp, suggesting a potential role by one of these domains in preventing proteasome-mediated degradation (167).

The PRO domain was expressed recombinantly in *Escherichia coli* in a 2012 paper that showed the enzyme was active against a common DUB fluorogenic substrate (Ub-AMC) and also was active against the two most-well characterized poly-Ub chain substrates (Lys48-linked, which is a marker for proteasomal degradation and Lys63-linked, which has more implications in signaling as it pertains to nutrient transport and plant development) (149, 169–171). The Ub pathway is highly regulated and has been characterized extensively and will be discussed in detail in future sections; however, findings that TYMV PRO can hydrolyze K48-poly-Ub chains bolstered former claims of TYMV PRO protection of the RdRp from proteasomal degradation (149, 167). Findings showed in *A. thaliana* protoplasts transfected with a plasmid encoding for Myc-tagged Ub, that the RdRp and a plasmids encoding variations of the PRO domain, following IP using an anti-RdRp antibody, RdRp ubiquitin conjugations were significantly decreased, showing that the RdRp is the target for PRO's DUB activity (149). A crystal structure of TYMV PRO was determined in 2013 and although it has two characteristic domains similar to papain and the active site being formed at the interface of these domains, the active site structurally has more homology to OTU DUBs than classic PLCPs, although OTU domains are technically considered

PLCPs (172). A 2020 publication revealed the crystal structure of TYMV PRO bound to Ub, which truly brought the story full circle (173), and this structure taken with the crystal structure of the enzyme alone will be discussed in length in future sections.

1.5 *Thesis Objectives*

Overall, the work presented focusses on the cysteine proteases of several different animal-, plant- and human-affecting +ssRNA viruses and their roles as it pertains to polyprotein processing and additionally to characterize not as commonly explored roles of these enzymes, such as programmed ribosomal frameshifting and deubiquitination. More specifically the goals are as outlined below:

1) A biochemical and biophysical investigation of the PRRSV PRF-stimulating complex of the nsp2 coding region: How PRRSV nsp1β and cellular PCBP2 bind the viral genome to induce PRF is still not very well understood. The research presented here sheds more light on this novel complex of a host protein-viral protein-viral genome interaction. Mutational analysis as well as biochemical and biophysical analysis of the complex was carried out for its characterization. Cell-based mutational assays also confirm biochemical assays and further show the role of nsp1β in innate immune suppression.
2) Exploring the polyprotein processing and deubiquitinating activity of marafivirus and benyvirus proteases: As outlined above, TYMV PRO has been extensively characterized, but the proteases from the closely related marafiviruses have not. The work presented shows that these enzymes have polyprotein cleavage tendencies that appear comparable to TYMV and also contain DUB activity to varying degrees. Studies on the benyvirus protease domains show interesting specifics on how these enzymes recognize Ub

substrates. Crystal structures of MRFV PRO provide evidence of the enzymes in the absence of presence of substrate, which shows the potentially dynamic nature of the enzyme.

This work has been done in relation to a bigger picture project of bioengineering transgenic plants that express selective ubiquitin-like inhibitors that would act as competitive inhibitors of these proteases, to ideally create some degree of viral resistance in the host. A similar strategy, with some modifications, is being used with the PLCP from SARS-CoV-2, which will be discussed herein.

2 *Chapter 2: Molecular characterization of the RNA-protein complex directing -2/-1 programmed ribosomal frameshifting during arterivirus replicase expression*

A majority of this research was originally published in the *Journal of Biological Chemistry.*

Patel, A., Treffers, E.E., Meier, M., Patel, T.R., Stetefeld, J., Snijder, E.J., Mark, B.L.. Molecular characterization of the RNA-protein complex directing -2/-1 programmed ribosomal frameshifting during arterivirus replicase expression. *J Biol Chem*. 2020. doi:10.1074/jbc.RA120.016105.

Contributions statement:

The conceptualization of this project was done prior to my arrival in the lab by B.L.M and E.J.S. Purification of all recombinantly expressed proteins and cloning of mutant proteins was done by myself. Electrophoretic mobility shift assays were done by me as well as purification of the tricomponent complex for all experiments. Analytical ultracentrifugation and data analysis was carried out by M.M. (University of Manitoba), Small-angle X-ray scattering data collection and processing was carried out by M.M. and T.R.P. (Lethbridge University), respectively. Cell culture experiments were carried out by E.E.T. Manuscript writing and compilation was primary done by myself and B.L.M, but all authors contributed to writing their experimental sections.

2.1 Introduction

RNA viruses have evolved remarkable noncanonical translational mechanisms to maximize the coding capacity of their genomes (10, 116), including the use of PRF. PRF enables the ribosome to access multiple overlapping ORFs within the viral genome (10, 174, 175), thus yielding alternative viral protein variants from what - upon cursory inspection - appears to be a single gene, allowing for the expression of partially colinear proteins with alternate C-terminal extensions and domains (10, 176, 177).

As discussed above, the first evidence for the occurrence of PRF was discovered in Rous sarcoma virus, which produces a gag-pol fusion protein from briefly overlapping gag and pol ORFs during infection (112, 113, 178). This is achieved by causing the host cell ribosome to slip back one position (-1 PRF) during translation of the viral RNA genome, which occurs at a heptameric 'slippery' sequence that is located 5-10 nucleotides upstream of an RNA structural element (stem-loop or pseudoknot) (112, 113, 178). Encountering this RNA structure causes the ribosome to pause and 'slip' on the slippery sequence, resulting in a -1 frameshift that opens access to an alternate reading frame (179). The frequency of frameshifting events differs per virus and presumably controls the stoichiometry of certain viral proteins (10).

Members of the order *Nidovirales* (including among others the families *Arteriviridae* and *Coronaviridae*) encode two large replicase polyproteins, pp1a and pp1ab, which are post-translationally cleaved to yield 12-16 mature nsps (28). Expression of pp1ab depends on a pseudoknot-stimulated -1 PRF event to occur in the short ORF1a/ORF1b overlap region (119, 176). Next to this well characterized -1 PRF event, most members of the arterivirus family also employ a more unusual -2 PRF mechanism. For example, in porcine reproductive and respiratory syndrome virus (PRRSV) and simian hemorrhagic fever virus (SHFV), -1 and -2 PRF events were

shown to occur at the same site in the nsp2-coding region of ORF1a, yielding two nsp2 variants. In the case of PRRSV these products are either truncated compared to full-length nsp2 (nsp2N, resulting from -1 PRF) or contain an alternative C-terminal domain (nsp2TF, resulting from -2 PRF) and were implicated in suppressing host innate immune responses (Fig. 2.1A) (89, 105, 131, 180). Interestingly, while a characteristic slippery sequence is present in the region of the PRRSV genome where these frameshifts occur, no discernible RNA secondary structural element could be predicted (105). However, a highly conserved C-rich motif (CCCANCUCC, or similar) is found 11 nt downstream of the slippery sequence shift in studied PRRSV isolates (Fig. 2.1B), which suggested that a novel transactivating mechanism facilitates PRF at this position as opposed to a ribosomal pausing mechanism that is usually induced by an RNA tertiary structural element (89, 123). Indeed, two *trans*-acting elements subsequently were shown to control -1 and -2 PRF in PRRSV: the PRRSV protein nsp1β and the host cell protein poly(C)-binding protein 1 or 2 (89, 105, 123). The viral and cellular proteins interact with each other and with the viral RNA genome to induce -1/-2 PRF in the nsp2-coding region of ORF1a. Although PCBP2 and nsp1β had previously been shown to interact with each other (124), the significance of this interaction for efficient PRF has only recently been discovered (89, 123).

PRRSV remains the most economically important viral disease in the swine industry (181) and its high pathogenicity may be due, in part, to the immune evasion mechanisms it employs during infection (28). Consequently, the further dissection of its molecular biology and gene expression mechanisms is highly relevant for efforts to improve PRRSV vaccines, including those based on attenuation by targeted engineering of the viral genome (182). Despite its importance to PRRSV replication, the biochemistry and structural biology of the interactions between nsp1β, PCBP1 or 2 and the PRRSV RNA genome have not been explored. Here we provide structural and

functional insights into the quaternary complex between nsp1β:PCBP2 and viral RNA that controls PRF. Site-directed mutations in both nsp1β and PCBP2 pinpointed key residues needed for complex formation with the RNA genome. Nsp1β mutagenesis was also used to identify residues essential to stimulate PRF as well as residues involved in the evasion of innate immune responses. While we found nsp1β and PCBP2 to be unstable on their own, combining the proteins with viral RNA containing the putative slippery sequence and C-rich motif resulted in a highly stable complex that we could study by analytical ultracentrifugation (AUC) and small-angle X-ray scattering (SAXS). Our study provides detailed molecular insights into a novel PRF-directing mechanism employing two protein transactivators interacting with the PRRSV genome to expand its coding capacity.

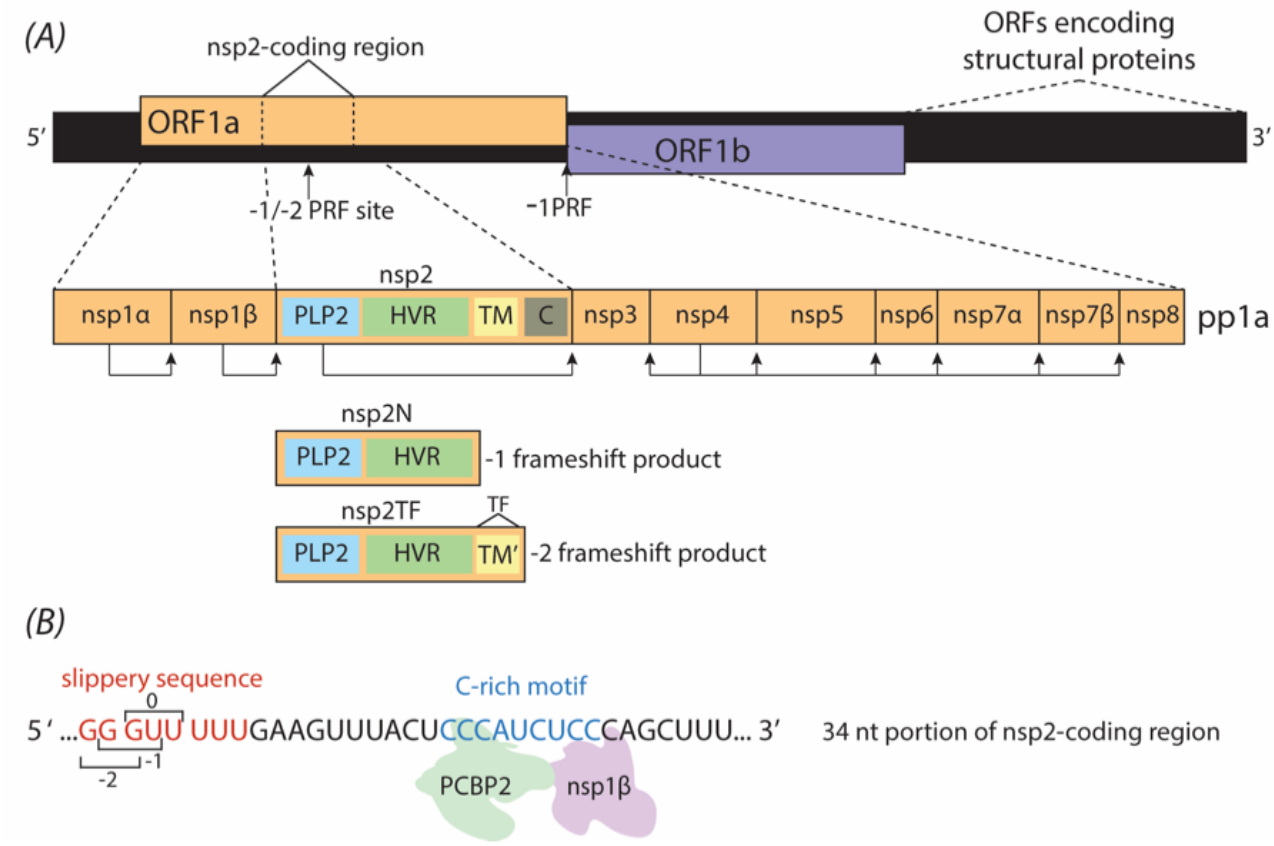


Figure 2.1 Simplified PRRSV genome, polyprotein and PRF events.

A) Schematic of the full +ssRNA PRRSV genome. Translation of the largest ORFs (ORF1a and ORF1b) yields replicase polyproteins 1a and 1ab (pp1a and pp1ab). Translation of ORF1b requires a -1 PRF event at the end of ORF1a. PRRSV pp1a is comprised of 10 nonstructural protein (nsp) subunits, four of which have autocatalytic polyprotein cleavage activity (arrows indicate cleavage sites). Whereas the –1 PRF at the ORF1a/1b junction is directed by stimulatory RNA structures, additional -1 and -2 PRF events occurring within the nsp2- coding region do not depend on higher-order RNA structures. These PRF events result in truncated nsp2 variants, nsp2N and nsp2TF, which both retain the papain-like cysteine protease (PLP2) and hypervariable regions (HVR) but lack the C-terminal Cys-rich domain (C). Nsp2TF also contains a modified transmembrane domain (TM') that is encoded by a short alternative ORF (TF) that overlaps with ORF1a in the -2 reading frame. B) The region of ORF1a from PRRSV SD01-08 showing where -1/-2 PRF occurs within nsp2. The slippery sequence is shown in red, and the C-rich motif is shown in blue. The C-rich motif replaces the canonical higher-order RNA structural element as found in most other PRF mechanisms and serves as a putative binding site for nsp1β and PCBP2. Adapted from (98, 99).

2.2 Materials and Methods

2.2.1 Primers used for DNA constructs

Table 2.1 Primers used for nsp1β and PCBP2 mutations for EMSAs

Oligo Name	Sequence (5'to 3')
nsp1β Y131A (5' Phos.) Forward	CCAACGCAGGCTTCAAGTCCG
nsp1β Y131A (5' Phos.) Reverse	AGCGCCTTACCATGCACGCCCC
nsp1β R135A (5' Phos.) Forward	CAAGTCCGCGGCATGCGTGC
nsp1β R135A (5' Phos.) Reverse	AGCGCGCGTTGGAGGTACTTAC
PCBP2 R40A Forward	GGAGAGTGGTGCACGTATCAAC
PCBP2 R40A Reverse	TCCGCCATCTTCTTAACTGATTC
PCBP2 R57A Forward	TATCACTTTGGCTGGACCCAC
PCBP2 R57A Reverse	ATCGCCTCAGGACAATTCCCTTC
PCBP2 N325D Forward	CAAAATTGCGGATCCAGTGGAAGG
PCBP2 N325D Reverse	ATCTGCGCCCCAGACATC

Table 2.2 Primers used for generating PCBP2 KH2 and KH1-KH2 proteins

Oligo Name	Sequence (5'to 3')
PCBP2 KH2 Forward	CTATAGGTCTCAAGGTGTCACCCTGAGGCTGGTGGTCC
PCBP2 KH2 Reverse	GATATAAAGCTTTTAAGTCTCCAACATGACCACGCAG
PCBP2 KH1-KH2 Forward	CTATAGGTCTCAAGGTAATGTCACTCTCACCATCCGGC
PCBP2 KH1-KH2 Reverse	GATATAAAGCTTTTAAGTCTCCAACATGACCACGCAG

Table 2.3 Primers used for frameshifting assays

Oligo Name	Sequence (5' to 3')	Purpose
EUnsp1b-G129A-fw	GGGCGTGCATGCTAAGTACCTCCAACGCAG	Site directed mutagenesis G129A
EUnsp1b-G129A-rev	TGGAGGTACTTAGCATGCACGCCCCACTTGG	Site directed mutagenesis G129A
EUnsp1b-K130A-fw	GGCGTGCATGGTGCUTACCTCCAACGCAG	Site directed mutagenesis K130A

EUnsp1b-K130A-rev	TGCGTTGGAGGTAAGCACCATGCACGCC	Site directed mutagenesis K130A
EUnsp1b-Y131A-fw	TGCATGGTAAGGCCCTCCAACGCAGGCTTC	Site directed mutagenesis Y131A
EUnsp1b-Y131A-rev	TGCGTTGGAGGGCCTTACCATGCACGCCCC	Site directed mutagenesis Y131A
EUnsp1b-L132A-fw	GCATGGTAAGTACGCCCAACGCAGGCTTCAAGTC	Site directed mutagenesis L132A
EUnsp1b-L132A-rev	GAAGCCTGCGTTGGGCGTACTTACCATGCACG	Site directed mutagenesis L132A
EUnsp1b-Q133A-fw	TAAGTACCTCGCACGCAGGCTTCAAGTCCG	Site directed mutagenesis Q133A
EUnsp1b-Q133A-rev	GAAGCCTGCGTGCGAGGTACTTACCATGCAC	Site directed mutagenesis Q133A
EUnsp1b-L136A-fw	CCAACGCAGGGCTCAAGTCCGCGGCATGC	Site directed mutagenesis L136A
EUnsp1b-L136A-rev	CATGCCGCGGACTTGAGCCCTGCGTTGGAG	Site directed mutagenesis L136A
EUnsp1b-Q137A-fw	TCCAACGCAGGCTTGCAGTCCGCGGCATGC	Site directed mutagenesis Q137A
EUnsp1b-Q137A-rev	CCGCGGACTGCAAGCCTGCGTTGGAGGTAC	Site directed mutagenesis Q137A
EUnsp1b-V138A-fw	GCAGGCTTCAAGCCCGCGGCATGCGTGCTG	Site directed mutagenesis V138A
EUnsp1b-V138A-rev	GCATGCCGCGGGCTTGAAGCCTGCGTTGGAG	Site directed mutagenesis V138A
EUnsp1b-R139A-fw	CTTCAAGTCGCTGGCATGCGTGCTGTGGTC	Site directed mutagenesis R139A
EUnsp1b-R139A-rev	CACAGCACGCATGCCAGCGACTTGAAGCC	Site directed mutagenesis R139A

EUnsp1b-G140A-fw	TTCAAGTCCGCGCCATGCGTGCTGTGGTCG	Site directed mutagenesis G140A
EUnsp1b-G140A-rev	CAGCACGCATGGCGCGGACTTGAAGCCTGC	Site directed mutagenesis G140A
EUnsp1b-M141A-fw	AGTCCGCGGCGCTCGTGCTGTGGTCGATCC	Site directed mutagenesis M141A
EUnsp1b-M141A-rev	CCACAGCACGAGCGCCGCGGACTTGAAGCC	Site directed mutagenesis M141A
EUnsp1b-R142A-fw	CCGCGGCATGGCTGCTGTGGTCGATCCTG	Site directed mutagenesis R142A
EUnsp1b-R142A-rev	CGACCACAGCAGCCATGCCGCGGACTTGAAG	Site directed mutagenesis R142A
EUnsp1b-fw	ACGAGAATTCCCATGGATGTCTGACGTTTACAGGTGGAAG	Amplify EUnsp1b and introduce upstream EcoRI and NcoI sites
3xFLAG-EUnsp1b-fw	GCCGAATTCCCATGGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGACGATGACAAGTCTGACGTTTACAGGTGGAAG	Amplify EUnsp1b and introduce upstream 3xFLAG tag and NcoI site
EUnsp1b-rev	ATCACCTGCAGGGCGGCCGCCTCGAGTCAGCCATACCACTTATGTG	Amplify EUnsp1b and introduce downstream XhoI, NotI and SbfI restriction sites
pL-seq	GATCTGATCTGGGGCCTC	Sequencing pL vector

2.2.2 Expression and purification of nsp1β (XH-GD)

Nsp1β from Chinese strain XH-GD, which is a strain of HP-PRRSV-2 was designed prior to my arrival to the lab using the exact same cloning strategy as described by Xue et al (183). The insert was placed into plasmid pGEX-6p-1 and was constructed by Blue Heron Biotech. The nsp1β

ORF is fused in-frame to glutathione S-transferase (GST) for affinity purification. This plasmid was transformed into *E. coli* BL21(DE3) Gold cells (Stratagene). The transformed cells were grown overnight at 37 °C in 25 ml lysogeny broth (LB) supplemented with 150 μg/ml ampicillin. The following morning, the culture was used to inoculate fresh LB with ampicillin (1:50 inoculation) and was subsequently grown at 37 °C to an A_{600} of 0.7-0.8. Expression of the GST-nsp1β (XH-GD) fusion protein was then induced by the addition of 0.5 mM isopropyl β-D-1 thiogalatopyranoside (IPTG) and left to incubate with shaking at 16 °C for an additional 18 h. Cells were then pelleted by centrifugation and stored at -80 °C.

Cell pellets were resuspended in ice-cold lysis buffer (50 mM 2-(*N*-morpholino)ethanesulfonic acid (MES) pH 6.0 and 500 mM NaCl) and lysed using a French pressure cell (AMINCO). Cell lysate was clarified by centrifugation (17,211 X g at 4 °C), and the supernatant containing the GST-nsp1β (XH-GD) fusion was mixed end-over-end for 1 h at 4 °C with GST-Bind resin (Millipore) that had been pre-equilibrated in lysis buffer. The lysate/resin slurry was poured into a gravity column and washed with 10 column volumes of lysis buffer, followed by elution of the fusion protein with lysis buffer supplemented with 10 mM reduced glutathione (adjusted to pH 6.0).

The GST tag was removed from nsp1β (XH-GD) using GST-tagged HRV 3C PreScission Protease, which was incubated with the eluted fusion protein in dialysis tubing overnight at 4 °C in 2 L of dialysis buffer (identical to lysis buffer). Tag-free nsp1β was separated from free GST and HRV 3C Precision Protease by passing the dialyzed protein mixture through GST-Bind resin (pre-equilibrated in dialysis buffer). The flow-through contained purified nsp1β and its concentration was quantified using a NanoDrop instrument (A_{280}, $\varepsilon/1000 = 40{,}910\ M^{-1}cm^{-1}$).

2.2.3 Expression and purification of nsp1β (SD01-08)

A pGEX-6p-2 expression vector of nsp1β from strain SD01-08, a North American PRRSV-2 strain, was kindly provided by Dr. Ian Brierley (University of Cambridge, Department of Pathology). The protein is expressed as a fusion protein in frame with N-term GST as described above. Expression was also carried out using *E. coli* BL21(DE3) Gold cells. Cells were grown overnight at 37 °C in 25 ml lysogeny broth (LB) containing 150 μg/ml ampicillin. The overnight culture was used to inoculate 1 L of fresh ampicillin-containing LB (1:50 inoculation) and was subsequently grown at 37 °C with shaking to an A_{600} of 0.7-0.8. Expression was induced using IPTG at a final concentration of 0.5 mM and cells were left to incubate with shaking at 16 °C for an additional 18 h. Cells were then pelleted by centrifugation and stored at -80 °C.

Cell pellets were resuspended in ice-cold lysis buffer (1X PBS (phosphate-buffered saline), pH 7.4, 300 mM NaCl, 100 mM KCl, 5 % glycerol, 2 mM DTT (dithiothreitol)) and lysed using a French pressure cell (AMINCO). Cell lysate was clarified by centrifugation (17,211 X g at 4 °C), and the supernatant containing the GST-nsp1β fusion was mixed end-over-end for 1 h at 4 °C with GST-Bind resin (Millipore) that had been pre-equilibrated in lysis buffer. The lysate/resin slurry was poured into a gravity column and washed with 10 column volumes of lysis buffer, followed by elution of the fusion protein with lysis buffer supplemented with 10 mM reduced glutathione (adjusted to pH 7.4).

The GST tag was removed from nsp1β using GST-tagged HRV 3C PreScission Protease, which was incubated with the eluted fusion protein in dialysis tubing overnight at 4 °C in 2 L of lysis buffer lacking additional NaCl. Tag-free nsp1β was separated from free GST and HRV 3C Precision Protease by passing the dialyzed protein mixture through GST-Bind resin (pre-

equilibrated in dialysis buffer). The flow-through contained purified nsp1β and its concentration was quantified using a NanoDrop instrument (A_{280}, ε/1000 = 23,786 $M^{-1}cm^{-1}$). Nsp1β (SD01-08) mutants Y131A and R135A were purified using the same method as described for the WT enzyme.

2.2.4 Expression and purification of PCBP2

Plasmid pQE-30-PCBP2 (Qiagen) encoding the full-length ORF for human PCBP2 with an in-frame 6xHis tag at its 5' end was provided by Dr. Ian Brierley (University of Cambridge, Department of Pathology). The plasmid was used to transform *E. coli* M15 (Qiagen). The transformed cells were grown overnight at 37 °C in LB broth containing both 35 μg/ml kanamycin and 150 μg/ml ampicillin. Subsequent culturing, IPTG-mediated induction of protein expression, cell lysis and lysate clarification were carried out as described above for nsp1β. Clarified lysate was mixed end-over-end for 1 h at 4 °C with nickel-nitrilotriacetic acid resin (Ni-NTA) (Qiagen) pre-equilibrated with lysis buffer. The lysate/resin slurry was then poured into a gravity column and washed with 10 column volumes of lysis buffer, followed by 10 column volumes of lysis buffer supplemented with 30 mM imidazole, and finally eluted with lysis buffer supplemented with 250 mM imidazole. The eluted protein was dialyzed against 2 L of buffer (1X PBS; pH 7.4, 100 mM KCl, 5 % glycerol) overnight at 4 °C and then further purified by gel filtration using a Superdex75 (GE healthcare) gel filtration column. The concentration of purified PCBP2 was quantified using a NanoDrop instrument (A_{280}, ε/1000 = 45,525 $M^{-1}cm^{-1}$). PCBP2 mutants (N325D; R40A, N325D; and R40A, R57A, N325D) were purified using the same method as described for the WT protein.

2.2.5 Differential scanning fluorimetry (DSF)

Purified nsp1β and PCBP2 were prepared as described above, however in a buffering system composed of 20 mM TRIS (Tris(Hydroxymethylamin-omethane)) pH 7.5 and 300 mM

NaCl. Commercial solubility and stability screens 1 & 2 from Hampton Research were used. 20 µl reactions were set up into Bio-Rad semi-skirted transparent 96-well PCR plates. Buffers/additives were added as per the manufacture's recommendations, protein was added to a final amount of 2 µg per well, Sypro Orange was added for a final concentration of 1X (Sigma-Aldrich) and was topped up to 20 µl with ultrapure water. Melt curves were generated on a Bio-Rad CFX96 qPCR machine with a gradient set from 4 °C to 95 °C with temperature increasing 1 °C per minute. Reads were taken on the fluorescence resonance energy transfer channel.

2.2.6 Cloning of PCBP2 KH2 and KH1-KH2 domains

PCBP2 KH2 and the KH1-KH2 fusion were cloned into the pE-SUMO expression vector (LifeSensors) using a pET-28a vector containing a PCBP2 insert as template for PCR, kindly provided by Dr. Ying Fang (Department of Pathobiology, College of Veterinary Medicine, University of Illinois, Urbana-Champaign, IL, USA). Primers that were used are indicated in Table 2.2. Amplicons generated had 5' BsaI and 3' HindIII restriction sites and were cloned into pE-SUMO using standard restriction cloning. This vector produces a fusion protein with an N-term 6xHis tag fused to SUMO (small ubiquitin-like modifier) to produce either 6xHis-SUMO-KH2 or 6xHIS-SUMO-KH1-KH2 with a SUMO protease cleavage site in between the SUMO and the protein of interest.

2.2.7 Expression and purification of PCBP2 KH2 and KH1-KH2

E. coli BL21(DE3) Gold cells transformed with pE-SUMO-KH2 or pE-SUMO-KH1-KH2 were grown overnight at 37 °C in LB broth containing 35 µg/ml kanamycin. Culturing and IPTG protein induction and expression was carried out identically to mentioned above for PCBP2 and nsp1β .

Cells were harvested after overnight expression and frozen at -80 °C or used immediately after harvesting. Cells were lysed and lysates were clarified as described above, however in a different lysis buffer (50 mM Bis-TRIS pH 6.8 (2-[Bis(2-hydroxyethyl)amino]-2-(hydroxymethyl)propane-1,3-diol), 0.5 M NaCl and 0.1 mM PMSF (phenylmethylsunfonyl fluoride). The lysate/resin slurry was then poured into a gravity column and washed with 10 column volumes of lysis buffer, followed by 10 column volumes of lysis buffer supplemented with 25 mM imidazole, and finally eluted with lysis buffer supplemented with 250 mM imidazole. In-house produced SUMO protease was added to the elution mixture to cleave off 6xHis-SUMO from either protein. The eluted protein with SUMO protease was dialyzed overnight at 4 °C against 2 L of dialysis buffer (50 mM Bis-TRIS pH 6.8 and 150 mM NaCl). The following day, the dialyzed sample was mixed with Ni-NTA pre-equilibrated with dialysis buffer to recapture any free 6xHis-SUMO for 1 h at 4 °C with end-over-end mixing. The flowthrough from this mixture contained purified KH2 or the KH1-KH2 fusion. The concentration of purified protein was quantified using Bradford assays, as neither protein contains Trp residues.

2.2.8 Site-directed mutagenesis

Nsp1β (SD01-08) variants used for EMSAs (Y131A and R135A) were constructed using round-the-horn site-directed mutagenesis (184) using plasmid pGEX-nsp1β WT as template (Table 2.1). The linear PCR amplicon was purified (Qiagen), followed by DpnI treatment to remove any plasmid template and then re-circularized using instant sticky-end DNA ligase (New England Biolabs). The ligation product was used to transform *E. coli* NEB5α (NEB). Once

confirming successful mutagenesis by DNA sequencing, plasmid pGEX-nsp1β Y131A/R135A were independently transformed into *E. coli* BL21 (DE3) Gold cells.

PCBP2 variants N325D; R40A, N325D; and R40A, R57A, N325D were constructed using a Q5 site-directed mutagenesis kit (New England Biolabs) and plasmid pQE-30-PCBP2 as template. Primers (Table 2.1) were designed using the NEBaseChanger tool to produce individual point mutations. Multiple rounds of site-directed mutagenesis were carried out to construct the double and triple PCBP2 mutations. All variants were confirmed by DNA sequencing.

Constructs used in cell culture assays were generated by standard PCR-based mutagenesis and recombinant DNA techniques. Expression vector pL1a was a derivative of an equine arteritis virus (EAV) ORF1a expression vector, in which the foreign gene is under control of a T7 RNA polymerase promoter and an encephalomyocarditis virus internal ribosomal entry site and is followed by a downstream T7 terminator sequence (185). pCAGGS-nsp1b-WT, R134A and R135A containing the European PRRSV strain SD01-08 nsp1β sequence were a kind gift from Dr. Ying Fang (Department of Pathobiology, College of Veterinary Medicine, University of Illinois, Urbana-Champaign, IL, USA). The wild type nsp1β (strain SD01-08) sequence was amplified using oligonucleotides EUnsp1b-fw and EUnsp1b-rev (all oligonucleotide sequences listed in Table 2.2) which introduced EcoRI and NcoI sites upstream and XhoI, NotI and SbfI sites downstream of nsp1β for cloning purposes. The PCR product was transferred to pUC19 vector for PCR-based mutagenesis to create mutants G129A, K130A, Y131A, L132A, Q133A, L136A, Q137A, V138A, R139A, G140A, M141A and R142A. The WT and mutant sequences were amplified using oligos 3xFLAG-EUnsp1b-fw and EUnsp1b-rev which introduced EcoRI and NcoI sites upstream of a 3xFLAG tag and transferred to the pL1a and pCAGGS (Addgene) expression vectors for cloning purposes. Correct introduction of the mutations was verified using Sanger

sequencing. pL-EUnsp2 (105), pL-EUnsp1β-2 (105), pLuc-IFN-β (186) and pcDNA-FLAG-MAVS are described as previously mentioned (187).

2.2.9 Electrophoretic mobility shift assays (EMSAs)

EMSAs were performed using synthetic ssRNA or ssDNA probes (Integrated DNA Technologies). Nsp1β and PCBP2 proteins used in the assays had been previously concentrated to 20 μM and frozen at – 80 °C in single-use aliquots. Each protein was thawed and diluted to 2 μM in EMSA reaction buffer (PBS pH 7.4, 100 mM KCl, 5% glycerol and 2 mM DTT). Nucleic acid probes were used at a final concentration of 20 nM. Protein(s) and nucleic acid were combined and co-incubated for each reaction with reaction buffer up to 20 μl for 10 min at 30 °C. Following incubation, each reaction was loaded onto a non-denaturing 8 % TBE-PAGE (TBE, Tris-Borate-Ethylenediamine tetraacetic acid; PAGE, Polyacrylamide gel electrophoresis) gel. Electrophoresis was performed for 70 min in ice-cold 0.5X TBE buffer at 140 V. The gel was subsequently stained with 1X SYBR gold nucleic acid stain (ThermoFisher Scientific) in 0.5X TBE for 30 min in the dark (to avoid photobleaching) prior to visualization using UV light.

2.2.10 Cell culture and antibodies

RK-13 (rat kidney) and HEK293T (human embryonic kidney) cells were cultured essentially as previously described (185, 188). mAb 58-46 (α-EU-nsp2) (105) which recognizes the N-terminal domain of nsp2, nsp2TF and nsp2N was a kind gift from Dr. Ying Fang (Department of Pathobiology, College of Veterinary Medicine, University of Illinois, Urbana-Champaign, IL, USA). mAb-FLAG (M2, Sigma).

2.2.11 Radioactive labeling and radio-immunoprecipitation analysis to determine frameshifting efficiencies

The frameshift stimulating abilities of the nsp1β mutants were determined by transient expression in RK-13 cells, using plasmid pL1a and the recombinant vaccinia virus/T7 polymerase expression system, which was performed essentially as described previously (185) by labeling transfected cells for 2 h using 150 μCi/mL of a (35S)Met/Cys mixture (EXPRE35S35SProtein Labeling Mix; Perkin-Elmer). Cells were transfected with nsp2 alone, nsp2 and wild type nsp1β expressed from the same plasmid or nsp2 and 3xFLAG-nsp1β wild type or mutants co-expressed from separate plasmids. Protocols for cell lysis, immunoprecipitation, SDS (sodium dodecyl sulfate)-PAGE, and quantification with a Typhoon Variable Mode Imager (GE Healthcare) have been described previously (98). Nsp2, nsp2TF and nsp2N were immunoprecipitated using mouse mAb 58-46 and the 3xFLAG-nsp1β mutants were immunoprecipitated using mouse mAb-FLAG (M2, Sigma). Band intensities (nsp2, nsp2TF and nsp2N) were quantified with ImageQuant TL (GE Healthcare) and normalized by the Met+Cys content of the respective products (nsp2: 14 Met, 32 Cys, nsp2TF: 14 Met, 24 Cys and nsp2N: 11 Met, 18 Cys) assuming that 35S Met and 35S Cys are incorporated with an efficiency ratio of 73:22 (the Met:Cys ratio in the mixture according to the manufacturer's documentation). We previously determined that calculated frameshifting efficiencies are only 1.06–1.07 times higher if equal incorporation efficiencies are assumed instead (105). Using these values, frameshifting efficiencies were calculated as (nsp2TF)/(nsp2 + nsp2TF + nsp2N) for −2 frameshifting, and (nsp2N)/(nsp2 + nsp2TF + nsp2N) for −1 frameshifting. The experiment was repeated three times.

2.2.12 Dual luciferase assay to determine interferon suppression

To determine the interferon suppression abilities of the nsp1β mutants 80% confluent HEK-293T cells in 24-well plates were transfected using the calcium phosphate transfection method (189). Cells were co-transfected with 5 ng pRL-TK, encoding *Renilla* luciferase (Promega), 25 ng pcDNA-FLAG-MAVS to induce an innate immune response, 50 ng pLuc-IFN-β, firefly reporter and 75 ng pCAGGS-3xFLAG-nsp1β expression plasmids. At 18 h post transfection, cells were harvested and luciferase expression was measured using the Dual Luciferase Stop & Glo Reporter Assay System (Promega) and the EnVision Multimode Microplate Reader (PerkinElmer). Experiments were performed in triplicate and independently repeated three times. Firefly luciferase activity was normalized by dividing the activity by the *Renilla* luciferase activity in the same well. Statistical significance ($p<0.001$) was determined using an unpaired two-tailed Student's t test in GraphPad Prism 8.1.1 (190).

2.2.13 Large-scale purification of the trimeric complexes for AUC and SAXS

Low concentrations of equimolar nsp1β and PCBP2 in PBS (pH 7.4), 100 mM KCl and 5% glycerol were co-incubated with a 1.1X molar excess of DNA/RNA probe for 3 h at 4 °C. Following incubation, the protein-nucleic acid complex was concentrated using a centrifugal filter unit (Amicon) to a volume of 2 ml. Any insoluble material was removed by high-speed centrifugation and the soluble material was loaded onto a Superdex200 (GE Healthcare) gel filtration column and purified. The integrity of purified complexes was evaluated by SDS-PAGE, non-denaturing TBE PAGE and dynamic light scattering prior to further analysis by analytical ultracentrifugation and small angle X-ray scattering.

2.2.14 Characterization of the trimeric complex by sedimentation velocity

Sedimentation velocity data were collected on a Beckman-Coulter ProteomeLab™ XL-I analytical ultracentrifuge equipped with an 8-hole An50Ti rotor. All samples were temperature equilibrated in the rotor for at least 2 h under vacuum.

Seven samples of 34 nt ssRNA at 8 μM (1x), 16 μM (2x), 32 μM(2x) and 64 μM (2x) concentrations were centrifuged at rotor speeds of 42000 rpm (samples ≥ 16 μM) or 30000 revolutions per minute (rpm) (8 μM) for 24 h at 20°C in the specified solvent. The concentration gradients in the cell were monitored by the absorbance optics at wavelengths of 278 nm (8 μM sample), 294 nm (samples ≥16 μM, 1st run) and 291 nm (samples ≥16 μM, 2nd run).

SEC-purified triple complex was prepared at 1.00 mg/ml, 2.00 mg/ml and 4.00 mg/ml total material concentration. Each sample was centrifuged at 30000 rpm for 24 h at 20°C in the specified solvent. The concentration gradients in the cells were monitored by the interference optics and by the absorbance optics at 300 nm wavelength. Due to its 1 order of magnitude higher absorption coefficient, the RNA dominated the signal in the absorbance optics. By contrast, the interference optics could detect all components. We repeated the experiment using the same samples in the same cells at 42000 rpm. Comparing the signal vs. loading concentrations of both runs, we detected that we had suffered material loss due to aggregation at the bottom of the cells. Actual sample concentrations in the second run had reduced to 0.44, 0.75 and 1.09 mg/ml.

The *c(s, f_r)* analysis (191) and *c(s)* analysis (192) were performed in SEDFIT. Direct fitting to the LAMM equation of sedimentation coefficient *s* and diffusion coefficient *D* of the observed species was executed in SEDPHAT using the *Global Discrete Species* model (193). In practice, due to the design of the software, we fitted $s_{20°C,w}$ and *M* using a substitute $\bar{\nu}$ of 0.73 and then

converted the values back to experimental s_e and D_e. 95% confidence intervals of the fitted parameters were determined in SEDPHAT (automatic confidence interval search with projection method (194–196). The partial specific volume $\bar{v}$ of the protein components were calculated using the program SEDNTERP 2 (197). The results were plotted using QTIPLOT(195), GUSSI (198) and MATPLOTLIB (199).

2.2.15 Measurement of the volumetric mass density of the solvent for sedimentation velocity

The density of the solvent was measured using a 1.000-1.220 floating hydrometer (Ertco, Germany) in a temperature controlled room. 6 independent measurements were taken. We measured three independent preparations of the buffer solution to assess variations between them. The confidence intervals in Appendix Table A1 include the deviations between the three preparations. The instrument yields the viscosity relative to pure water (ρ_r) and the readings were converted to absolute density (ρ), using the volumetric mass density of water (ρ_w = 0.998234 g/cm^3).

$$\rho = \rho_r * \rho_w$$

2.2.16 Solvent viscosity measurements for sedimentation velocity.

The viscosity of the solvent was measured using an SV-10 tuning fork vibro viscometer (A&D Company, Missisauga, Canada) in a temperature controlled room set to 19 ± 1 °C. The sample cup was filled with 40.0 ml ultrapure water and the pedestal height adjusted until the water surface reached the tapered region on the sensor plates. The instrument was then calibrated with this water using the "Simplified Calibration" function. The water was carefully drained from the

sample cup using a syringe with an attached flexible tube without disturbing the geometry of the setup (such as lowering the pedestal or moving the sample cup or shifting the sensor protector). 40.0 ml solvent with a temperature of approximate 21 °C was then added to the sample cup (without disturbing the geometry) and let to cool while continuously monitoring the viscosity every 15 s using the RsVisco control program. All measurement values from 20.1 - 19.9 °C were selected and averaged. The calibration/measurement cycle was repeated 6 times, yielding 6 independent measurements. We measured three independent preparations of the buffer solution to assess variations between them. The confidence intervals in Appendix Table A1 include the deviations between the three preparations. The instrument yields the viscosity relative to pure water (η_r) and had to be converted to absolute viscosity (η).

$$\eta = \eta_r / \rho_r$$

2.2.17 SAXS data collection and processing

The SAXS data for nsp1β:PCBP2:ssRNA complex was collected and processed as described previously (200, 201) at the B21 beamline at Diamond Light Source (Didcot, Oxfordshire, UK). Briefly, 50 µL of the 10 mg/ml complex was injected into 4.5 ml Shodex KW40 size exclusion column connected to an in-line Agilent 1200 (Agilent Technologies, Stockport, UK) HPLC, a flow cell and an Eiger 4M X-ray detector. We collected ~600 frames where each frame was exposed to the X-rays for 3 seconds. The peak region was buffer subtracted and merged using Primus (202), followed by Guinier analysis of merged data. Dimensionless Kratky analysis was also performed to ensure that the complex is folded. The pair-distance distribution ($P(r)$) analysis was performed using the program GNOM (203) to obtain the R_g and D_{max}. Next, we calculated 20 low-resolution structures using the $P(r)$ information and program DAMMIN (204). Lastly, the 20

low-resolution structures were averaged and filtered to obtain a representative structure using the DAMAVER package (205), as described previously (200).

2.2.18 SAXS envelope fitting

The experimentally determined envelope was used to manually fit in pre-existing protein structures and a modeled nucleic acid molecule representing the 34-nt ssRNA. The crystal structure of nsp1β from PRRSV strain XH-GD (PDB ID: 3MTV (183); purple with the RBM helix in red) was fit in as a monomer. The NMR solution structure of the KH1-KH2 fusion was also fit in (PDB ID: 2JZX (206); yellow and pink, respectively). The third KH domain was fit in from an existing crystal structure bound to C-rich DNA with the nucleic acid removed (PDB ID: 2P2R; teal). RNA in green is from a structure of KH1 bound to C-rich RNA as a reference (PDB ID: 2PY9 (207)). The 34 nt ssRNA molecule (orange with C-rich motif shown in blue) was modeled using w3DNA 2.0 (208) and subsequently fit into the experimentally determined density. All fitting was completed in PyMOL (209).

2.3 Results and Discussion

2.3.1 Determination of constructs for nsp1β and PCBP2 purification

Originally, nsp1β from an HP-PRRSV strain termed XH-GD was used, as a crystal structure for the protein was determined in 2010 (183). The DNA construct was cloned exactly as described by the authors and the same was done for expression and purification. The buffering system the authors used for this protein was MES pH 6.0 with 0.5 M NaCl and crystals were obtained after concentrating the protein to ~3.5 mg/ml. The fact that this protein was able to be crystalized made us believe that the protein was relatively stable and soluble at these elevated concentrations thusly making us want to proceed with this version of nsp1β as the end goal of this project originally was to structurally characterize the tricomponent complex through crystallization, which undoubtably relies on moderately high protein concentrations. Unfortunately, in our hands, after many attempts and experimental modifications, this protein was not as soluble and stable as initially thought and was never able to be concentrated higher than ~0.8 mg/ml until severe precipitation, though was able to be purified at lower concentrations (Fig. 2.2A). However, now knowing this, an alternate construct of nsp1β from a PRRSV-2 strain termed SD01-08 (~45 % aa sequence identity) was proceeded with. This version of the protein was able to be purified to roughly the same concentrations as XH-GD and our collaborators more frequently use constructs and a reverse genetics system based on this strain, which would ultimately be more representative and complementary to our work (Fig. 2.2E).

PCBP2 was also expressed and purified in-house (Fig. 2.2B) and also proved to be a challenging protein to work with. The protein behaved very similar to both versions of nsp1β, not being able to concentrate it alone much past ~1 mg/ml, which was again cause for concern for crystallographic studies. In looking into the literature, no full-length structure of PCBP2 has been

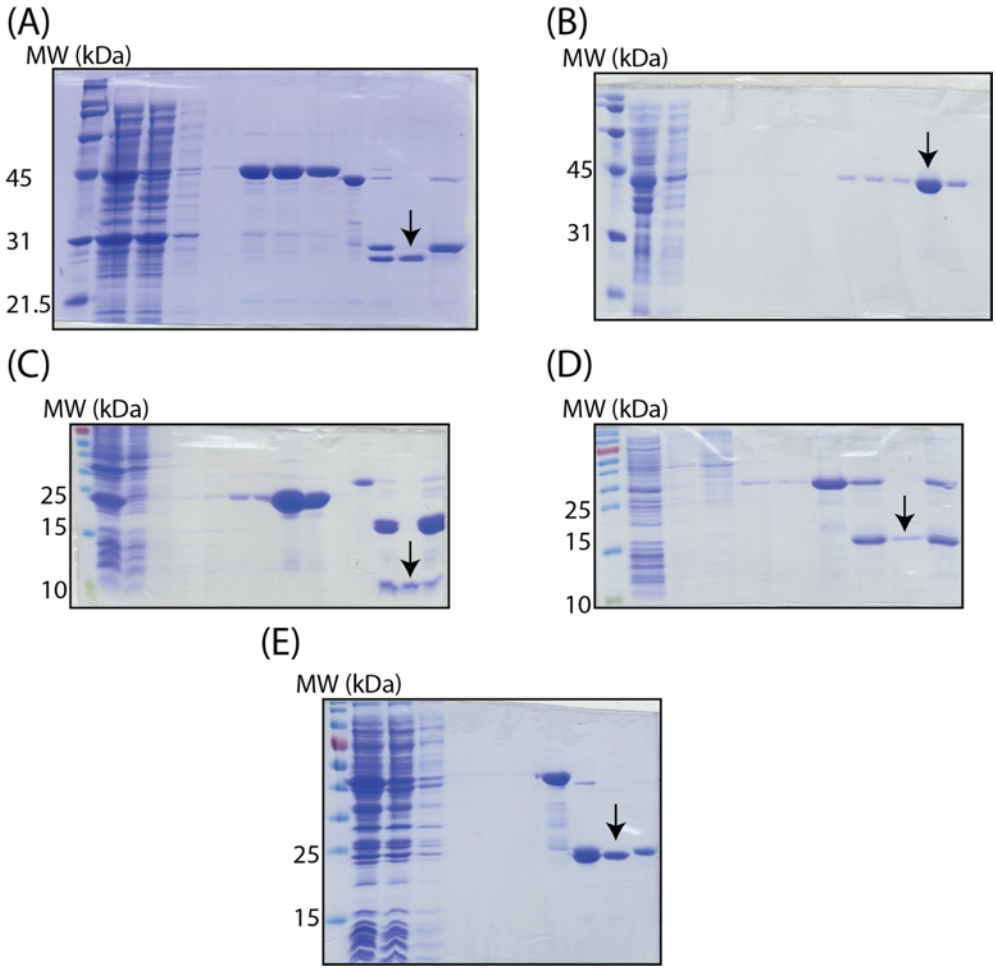


Figure 2.2 Purification of nsp1β, PCBP2 and KH domains.

A) Purification of PRRSV XH-GD nsp1β (23.5 kDa). B) SDS-PAGE gel of the purification of PCBP2 (38.5 kDa). C) SDS-PAGE gels of the purification of the KH2 domain of PCBP2 (7.5 kDa) and a KH1-KH2 fusion (16.9 kDa) (D). E) SDS-PAGE gel of the purification of PRRSV SD01-08 nsp1β (23.4 kDa). Arrows indicate final purified protein. First lane in all gels depict ladder, followed by clarified lysate, flow through and washing steps leading up to final purified proteins (arrow).

deposited into the PDB, however crystal and nuclear magnetic resonance (NMR) structures of the individual KH domains have been determined (206, 207, 210, 211), leading us to design constructs that generated either the KH2 domain or a KH1-KH2 fusion, as we believed that purifying the individual domains, which are quite compact, would be more amenable towards our studies (Fig 2.2C & D). Previous studies have found that the individual domains are enough to bind C-rich nucleic acid and alone can crystalize in complex with nucleic acid (127, 210, 211). Further, a yeast two-hybrid system suggested that the KH2 domain of PCBP2 is essential in interacting with nsp1β, focussing more of our attention to this domain (125). However, not even seven months into my degree, we were given insight from a paper that was subsequently published in 2016 that both the KH1 and KH3 domains seem to be involved with stimulating PRF (123), essentially making it more logical that we work with the full-length protein as the KH1 and KH3 domains are located at the N- and C-term of PCBP2, respectively.

2.3.2 Optimizing buffering conditions for nsp1β and PCBP2

Now knowing which constructs would have to be used for our characterization of the tricomponent complex, DSF was carried out in order to optimize buffering conditions for both nsp1β (SD01-08) and full-length PCBP2. It was important to optimize these conditions for both proteins and attempt to find a condition that satisfied both proteins for subsequent studies. This was done using commercial screens as described above and SYPRO orange. The dye binds to the hydrophobic core of folded proteins, quenching the fluorescence of the dye. Upon applying a linear, increasing temperature gradient to a protein in 96 different conditions using a qPCR machine, the protein will unfold, releasing the dye, which gives off detectable fluorescence and can be correlated to the protein's melting temperature (212). The first screen that was used was

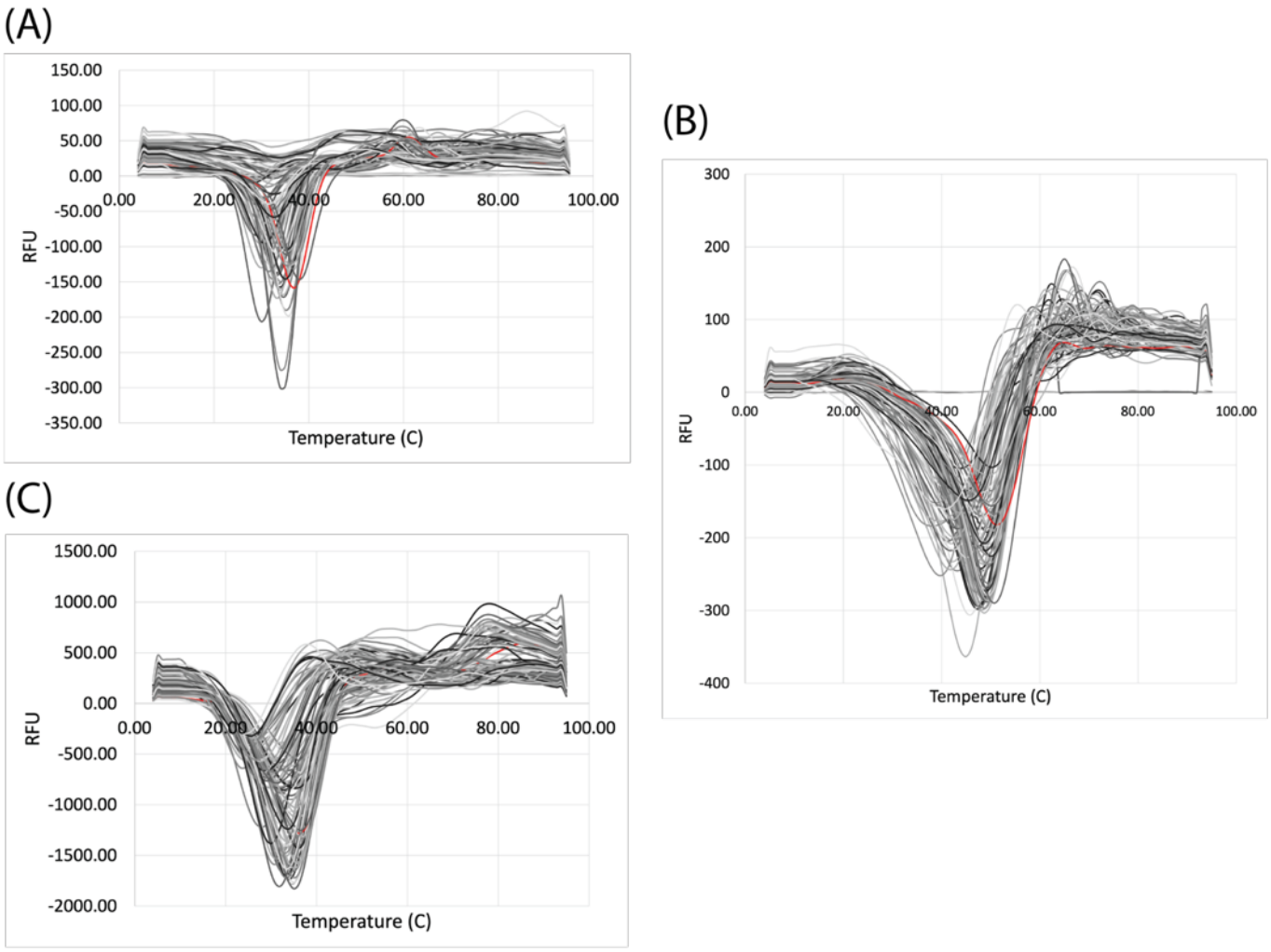


Figure 2.3 Differential scanning fluorimetry with nsp1β and PCBP2.

A) DSF performed on nsp1β (SD01-08) to determine suitable buffering system where red line indicates PBS. B) DSF performed on PCBP2 where red line indicates PBS. C) DSF performed on nsp1β to determine additives that increase melting temperature where the hidden red line indicates KCl.

the solubility and stability screen 2 (Hampton Research), which focusses on buffering system optimization from pH 4.5 sodium acetate to pH 9.5 glycine, increasing in 0.5 units of pH stepwise with a variety of different buffers, however, one quarter of the control wells (2 of 8) were altered to include PBS (pH 7.4), as all structures deposited of the individual KH domains were determined in a final buffering system including PBS.

For nsp1β, the melting temperature in the initial TRIS pH 7.5 buffer was ~32 °C. The highest temperature was achieved in a BIS-TRIS propane buffer (pH 8.5) at ~39 °C, however pH 7.4 PBS (Fig. 2.3A, red line) was close behind, increasing the initial melting temperature by 6 °C to a final of 38 °C. PCBP2 seemed to be much more stable in the same initial TRIS buffer with a melting temperature of ~44 °C, however the PBS (Fig. 2.3B, red line) and a glycine buffer (pH 9.5) seemed to impart the highest degree of stability, increasing the melting temperature to ~51 °C. Since PBS seemed to be the buffer that imparted the highest shared melting temperature, it was selected for subsequent experimentation. However, the relatively low melting temperature of nsp1β in comparison to PCBP2 was something that could be further optimized with the incorporation of additives. A similar assay was performed with 96 different additives using nsp1β in the initial TRIS buffer (Fig 2.3C). Although two zwitterionic compounds seemed to have the biggest impact on increasing melting temperatures, the simpler KCl seemed to have comparable effects, increasing the melting temperature from 32 °C to 38 °C (Fig 2.3C, red line) and we presumed would be less impactful in downstream studies and thusly was selected as a suitable additive to the PBS buffer, which contained only low amounts of KCl originally.

2.3.3 EMSAs using 54 nt RNA and DNA probes

EMSAs were first performed using a stretch of nucleic acid representing the PRRSV SD01-08 RNA genome that was 54 nt in length, including the slippery sequence, C-rich motif, and a length 3' tail (Fig 2.4A). The analogous DNA probe (uracil to thymidine) was also used to finetune the system, as it was much more cost-effective and less fastidious to work with. Structural studies indicate that the methyl group of thymine, which is not present on uracil, does not interact with the KH domains of PCBP2, and that it is the O2 and N3 groups of a thymine/uracil nucleoside that interact with the PCBP2 amide backbone directly (207, 210). Nucleic acid (where applicable) was used at a final concentration of 20 nM and nsp1β and PCBP2 protein concentration was used identically in an 8X to 30X molar excess concentration to probe. For both DNA and RNA (Fig 2.4 B & C, respectively) either protein alone does not seem to interact with the nucleic acid probe, however, a gel shift was visible upon addition of both proteins in tandem. Similar trends in protein-nucleic acid binding were seen for DNA as well as the more representative RNA, which was promising to see, as this validated our strategy in truncating our probe using proxy DNA oligos in lieu of RNA.

2.3.4 Probe truncation

Crystallization of this complex was the ultimate goal of this chapter, so it was important to truncate the nucleic acid to as small as possible and still have a well-behaved complex that bound the probe. This was important as unbound nucleic acid would be quite flexible and would prove challenging in crystallization, as it could ultimately affect crystal packing. The probe size was systematically truncated from 54 nt to 34 nt using DNA oligos (Fig. 2.4A). Bases were removed

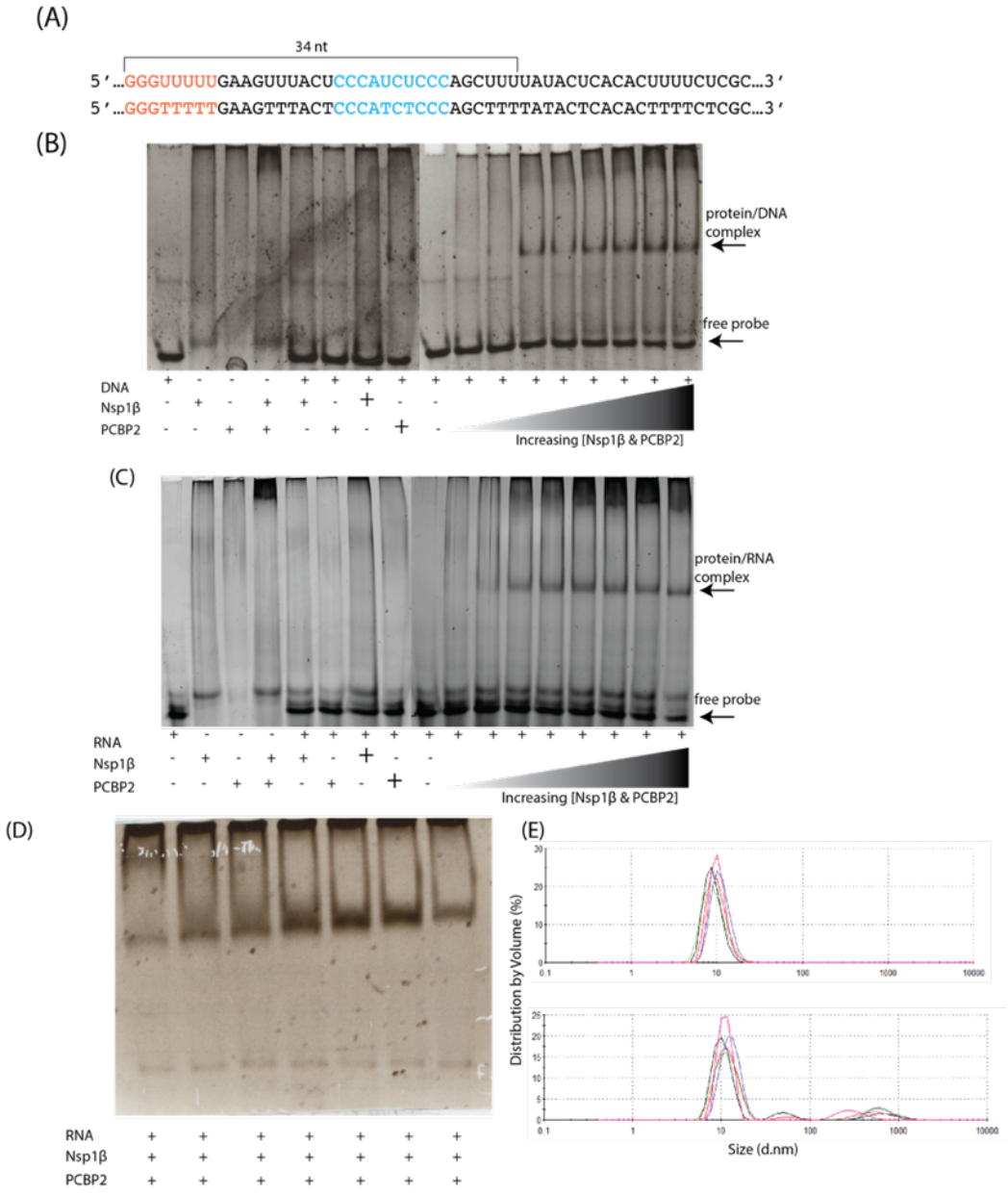


Figure 2.4 EMSAs for truncating the probe size and DLS on the purified complex.

A) Original 54 ntRNA and ssDNA probes corresponding to a stretch of the nsp2-coding region of the PRRSV SD-0108 ssRNA genome where PRF occurs. B) and C) EMSAs performed with 20 nM of a 54-nt ssDNA and RNA probe. Nsp1β and PCBP2 were combined independently or in tandem with the nucleic acid probes. D) Systematic truncation of the 54 nt probe to 30 nt using DNA as a proxy. E) DLS data of the purified complexes (34 nt, top; 30 nt, bottom) collected at ambient temperature at 2 mg/ml and represented as sample percent distribution by volume as a function of particle size.

from the 3' end of the initial probe as it was hypothesized that the slippery sequence and C-rich motif may be essential in complex formation, as well as potentially the linker between the two motifs. The complex still formed down to 30 nt (final lane Fig 2.4D) and even slightly smaller; however dynamic light scattering (DLS) data on the purified tricomponent complex indicated a lack of sample monodispersity and potential aggregation upon truncations smaller than 34 nt (Fig 2.4E 34 nt (top) and 30 nt (bottom)), thusly 34 nt was selected as the final probe size.

2.3.5 Formation of the most stable and compact quaternary complexes

To characterize the biochemistry of this complex in greater detail, we carried out a series of EMSAs as above mentioned with nucleic acid probes of systematically decreasing size to identify the shortest RNA fragment to which the proteins would stably bind, with the aim of identifying a compact protein:RNA complex amenable to preparative (milligram) scale purification. Using the 34-nt probe, an EMSA was initially performed with WT nsp1β, PCBP2 and the ssDNA nucleic acid probe. As shown in Fig. 2.5B & C, nsp1β does not appear to interact with the nucleic acid alone even at a 20X-fold molar excess in relationship to the probe, which is consistent with previous findings (123). Interestingly, with this shortened probe, PCBP2 does interact with the ssDNA probe on its own (Fig. 2.5B), but not with ssRNA (Fig. 2.5C). This can be seen as low as an 8X- fold molar excess but is highly amplified when the amount of PCBP2 is increased, as seen when a 20X-fold molar excess is added in relation to nucleic acid. When both nsp1β and PCBP2 are present with the probe, a shift can be seen in comparison to PCBP2 bound to DNA alone, indicating the formation of a trimeric complex.

Lastly, we wanted to confirm the importance of the cytosine-rich motif as it pertains to

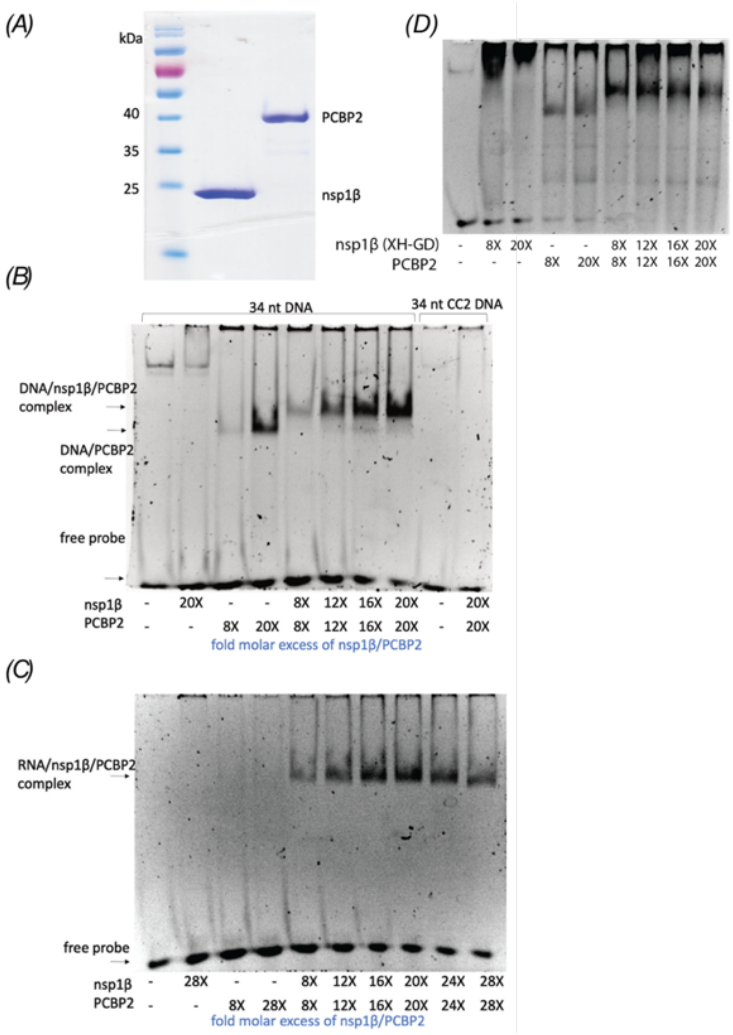


Figure 2.5 Purification of Nsp1b and PCBP2 and their interaction with the 34 nt DNA and RNA probes.

(A) SDS-PAGE showing recombinant PRRSV SD01-08 nsp1β (~23 kDa) and human PCBP2 (~38 kDa) in final PBS purification buffer. *(B)* Electrophoretic mobility shift assay (EMSA) performed with 20 μM of a 34-nt ssDNA probe corresponding to a stretch of the nsp2-coding region of the PRRSV SD01-08 ssRNA genome where PRF occurs. Nsp1β and PCBP2 were combined independently or in tandem with the nucleic acid probe. The molar excess of each protein relative to the nucleic acid probe is shown below each lane. Lanes 9 and 10 contain a control DNA probe (CC2) (105) in which the C-rich region has been altered to adenine/guanine nucleobases. *(C)* EMSA performed with 20 μM of a 34-nt ssRNA probe identical to a stretch of the nsp2-coding region of the PRRSV ssRNA genome where PRF occurs. Nsp1β and PCBP2 were combined independently or in tandem. (*D)* Similar EMSA performed as in *(B)*, but with nsp1β XH-GD. The molar excess of each protein relative to the RNA probe is shown below each well.

complex formation. The “CCCATCTCC” stretch of the ssDNA probe was mutated to “GAAATATGG”, which is termed the 34-nt CC2 DNA (105). This probe had all cytosine nucleobases of the C-rich motif altered to guanines or adenines to see if complex formation, as it pertains to this specific probe, could be abolished simply by disrupting this previously implicated binding site (89, 105, 123). As it can be seen, even in the presence of a large molar excess of both proteins, no complex formation is detectable with the mutated probe, indicating the crucial role of cytosine or potentially CT repeats. These repeats have been implicated in being present in ssDNA sequences that interact with the KH domains of PCBP2 (126, 127, 213). It can also be theorized that base stacking within CT repeats may arise in the nucleic acid molecule further lends itself towards these nucleobase-amino acid interactions (126, 206, 211). Results using DNA (Fig. 2.5B) suggest that PCBP2 recruits nsp1β to the nucleic acid; however, when the assay was performed with a 34-nt ssRNA probe (Fig. 2.5C), even with a large molar excess of PCBP2 (28X), PCBP2 did not appreciably interact with the RNA probe alone. This is consistent with previous work demonstrating that PCBP1 and 2 have a higher affinity towards ssDNA compared to ssRNA (213, 214). Our results and those of others (123) also demonstrate that nsp1β does not interact with RNA or DNA independently. Thus, nsp1β may bind PCBP2 to enhance its affinity for RNA and thereby enable all three components to assemble into a quaternary complex that promotes frameshifting. Indeed, as we found for the ssDNA probe, when nsp1β, PCBP2 and ssRNA are combined in tandem, they form a readily detectable complex *in vitro* (Fig. 2.5C). A similar EMSA was performed using nsp1β XH-GD (Fig 2.5D) that was purified in the PBS buffering system and showed analogous results as nsp1β SD01-08 in which it does not bind the probe alone, however, in the presence of PCBP2, a shift is seen indicative of the tricomponent complex, showing that both versions of nsp1β behave similarly as it pertains to complex formation.

2.3.6 Probing the protein-RNA binding interface of the PCBP2:nsp1β:RNA complex

Guided by known NMR and X-ray structures of PCBP2 KH domains (210, 211) and full-length nsp1β (183), we generated a number of site-directed mutations aimed at identifying residues of PCBP2 and nsp1β that are essential for RNA binding and PRF. For PCBP2, three variants were made within the first and third KH domains (Fig. 2.6A), including a single mutant (N325D), a double mutant (N325D; R40A) and a triple mutant (N325D; R40A; R57A). The N325D single mutation was constructed on the basis of a crystal structure of the KH3 domain bound to a short fragment of C-rich, ssDNA with sequence AACCCTA (PDB ID: 2P2R). As shown in Fig. 2.6A, the carboxamide sidechain of N325 is within hydrogen bonding distance (3.2 Å) of N1 of an adenine base. The N325D mutation retains the overall stereochemistry of asparagine, but imparts a negative charge that we predicted would disrupt nucleic acid binding without altering the structure of the KH domain. The R40A and R57A mutations were constructed based on the crystal structure of the KH1 domain bound to C-rich ssDNA (AACCCTAACCCT) (PDB ID: 2PQU). The guanidinium group of Arg40 forms two interactions with the keto group of a cytosine nucleobase by the formation of hydrogen bonds of 3.1 and 2.9 Å (Fig. 2.6A). This interaction suggests Arg40 may be a key residue for interaction with C-rich motif of the PRRSV genome. Similarly, the guanidinium group of Arg57 forms a 3.2 Å hydrogen bond with an additional cytosine base that we also predicted participates in binding the C-rich motif (Fig. 2.1C and 2.6A). Using these mutations, EMSAs were first carried out using WT nsp1β to gain insight into how the PCPB2 mutations affected complex formation (Fig. 2.6C). In comparison to WT PCBP2, the N325D

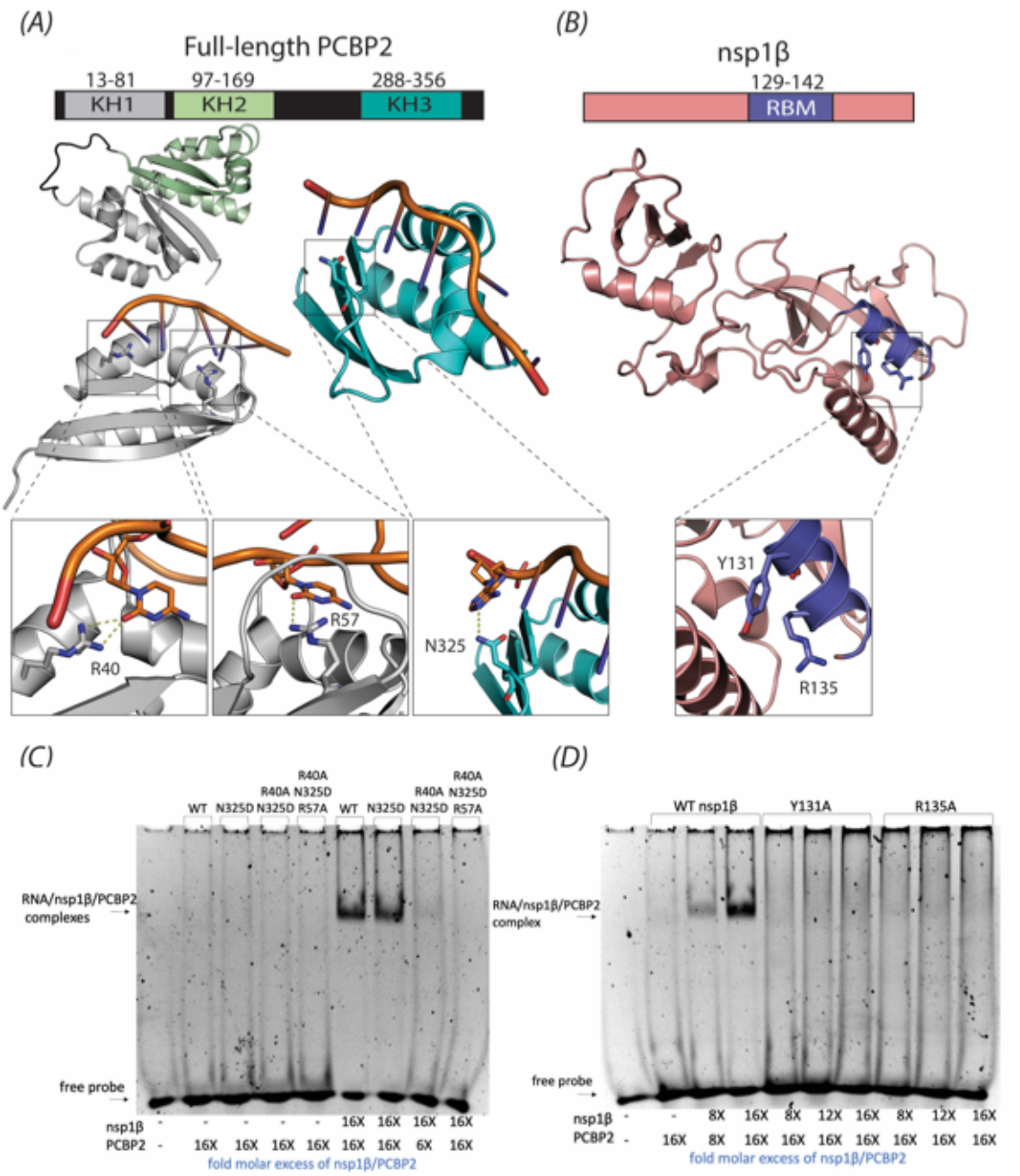


Figure 2.6 Structure-guided mutational analysis of PCBP2 and nsp1β binding to PRRSV RNA.

(A) Probing nucleic acid interaction sites of PCBP2. A schematic of full-length PCBP2 showing KH1 (grey), KH2 (green) and KH3 (teal) domains with accompanying 3-dimensional structures (PDB IDs: 2P2R (211), 2JZX (206) and 2PQU (210) for DNA-bound KH1, KH1-KH2 fusion and DNA-bound KH3, respectively). The KH1 guanidinio groups of Arg40 and Arg57 appear to hydrogen-bond with the keto group of a cytosine nucleobase, while the sidechain of Asn325 in KH3 is within hydrogen bonding distance of an adenine nucleobase. While the published structure has the amino group of N325 interacting with adenine, it is more likely that the carboxamide is

rotated 180° to allow the carbonyl group to interact with the base instead. *(B)* Probing PRRSV nsp1β interactions with nucleic acid. A schematic of nsp1β from PRRSV strain XH-GD (PDB ID: 3MTV (92)) with the putative RNA-binding motif (105) and residues Tyr131 and Arg135 shown in purple. Figures were generated using PyMOL (209). *(C)* and *(D)* EMSAs performed with 20 μM of the 34-nt ssRNA (Fig 1B). In *(C)*, WT nsp1β and two mutants (Y131A and R135A) were combined with PCBP2 and the ssRNA probe. In *(D)*, WT PCBP2 and three mutants (single mutant (N325D), double mutant (N325D; R40A) and triple mutant (N325D; R40A; R57A) were combined with nsp1β and the ssRNA probe. Molar excess of each protein is listed below each well in comparison to the probe.

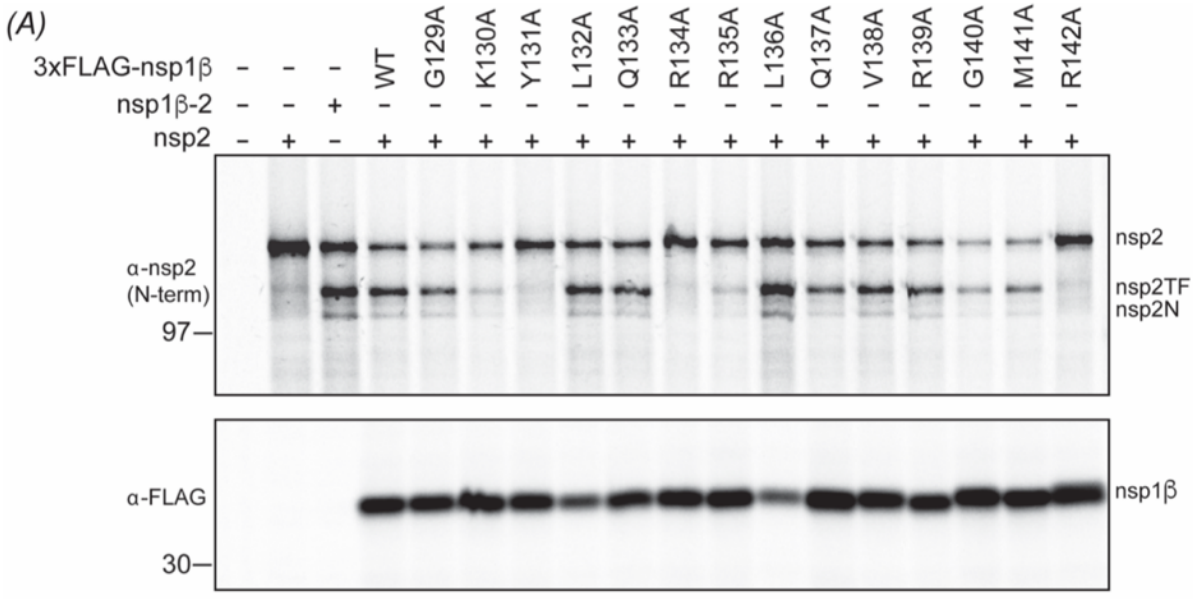
mutation alone did not affect complex formation. Given the significant electrostatic repulsion that was predicted to occur, this finding suggests a lesser role for KH3 in binding to the C-rich motif in the PRRSV genome. In contrast, when the double mutant R40A + N325D was assayed, a marked decrease in complex formation was observed, implicating Arg40 as a key player in binding the C-rich motif. The PCBP2 triple mutant (R40A, N325D, R57A), in which two mutations were made in KH1 and one in KH3 domains, abolished PCBP2's binding capabilities to the probe and subsequently complex formation.

To probe nsp1β residues that are crucial for complex formation, the proposed RNA-binding motif (RBM, Fig. 2.6B) that is highly conserved within almost all PRRSV isolates (215) was analyzed by site-directed mutagenesis. This motif with the sequence GKYLQRRLQ is comprised of several basic amino acids that have been implicated in -2/-1 PRF stimulation (89), innate immune suppression (109, 215), and nuclear polyA mRNA retention of host cell transcripts, which prevents cytoplasmic entry and subsequent translation of essential cellular mRNAs in PRRSV-infected cells (109). Previous studies have shown that mutations within the RBM decreased the prevalence of frameshifting products nsp2TF and nsp2N (89) and may limit the ability of PRRSV to suppress the host innate immune response (215). Indeed, it has been speculated that the nsp2TF and nsp2N frameshifting products aid in suppressing the innate immune response (131). Nsp2 has an N-terminal papain-like cysteine protease domain (PLP2 in Fig. 2.1A) that functions in viral replicase polyprotein processing, but also has deubiquitinating and de-ISGylating activities that are thought to help the virus evade porcine immunity pathways (132, 216). It stands to reason then that these auxiliary functions of nsp2 are heightened with the translation of the nsp2-variant frameshifting products due to the presence of the PLP2 domain within all three proteins (131). Previous studies revealed that nsp2 and nsp2TF are both membrane-associated, but are targeted to

different compartments in the infected cell (105). Furthermore, nsp2N lacks a predicted transmembrane domain (Fig. 2.1A) that would tether it to a membrane. It may thus be a cytosolic protein (131), possibly acting as a deubiquitinase that corrupts the host ubiquitin system to suppress innate immune responses (217). To investigate the biochemistry of the nsp1β RBM motif, two point mutations (Y131A and R135A) were independently introduced. In contrast to WT nsp1β, the nsp1β-Y131A and nsp1β-R135A mutations abolished complex formation with RNA when combined with WT PCBP2 (Fig. 2.6D), implicating these residues in the formation of the quaternary complex. Given these results, and previous mutational analyses of the region (123), the nsp1β RBM promotes RNA binding but only in the presence of PCBP2. Previous yeast two-hybrid experiments have found the two proteins to interact (218), suggesting that their binding may induce conformational changes in one or both proteins that favour RNA binding and stimulation of -1/-2 PRF, since neither appreciably binds RNA on its own (Fig 2.5C and D).

2.3.7 Mutations in the nsp1β RBM motif also prevent -1/-2 PRF product formation

To gain deeper insight into the role of the nsp1β RBM motif in -1/-2 PRF stimulation, we systematically analyzed its role in -1/-2 PRF by mutating each residue to an alanine (G129 to R142 of the SD01-08 PRRSV strain). The nsp1β expression plasmids were co-transfected with a plasmid expressing SD01-08 nsp2 into RK13 cells that were infected with a recombinant vaccinia virus expressing T7 RNA polymerase. Subsequently, nsp2, nsp2TF and nsp2N were metabolically labeled, immunoprecipitated and separated by SDS-PAGE. The expression of the three nsp2 variants was quantified in each condition and compared to the situation in which nsp2 was co-expressed with WT nsp1β. Nsp1β mutant expression was confirmed by immunoprecipitation using an antiserum recognizing the N-terminal 3xFLAG tag.



(B)

	nsp2	nsp1b-2	WT	G129A	K130A	Y131A	L132A	Q133A	R134A	R135A	L136A	Q137A	V138A	R139A	G140A	M141A	R142A
% nsp2 (no PRF)	95	43	35	38	79	94	39	43	95	86	35	58	38	40	39	36	96
% nsp2TF (-2 PRF)	4	43	56	52	18	5	51	49	4	11	53	38	53	53	51	56	3
% nsp2N (-1 PRF)	1	13	9	10	4	1	9	8	0	3	12	4	9	7	10	8	1

Figure 2.7 Analysis of trans-activating frameshift stimulation by nsp1β mutants.

Plasmids expressing SD01-08 PRRSV WT or mutant nsp1β were co-transfected with a plasmid expressing nsp2 in RK13 cells infected with a recombinant vaccinia virus expressing T7 RNA polymerase. As controls single expression of nsp2 or nsp1β, expression of a self-cleaving nsp1β-2 polyprotein, and a nontransfected sample were included. (*A)* Following ^{35}S metabolic labeling, proteins were immunoprecipitated with mAb58-46 (nsp2, nsp2TF and nsp2N) or mAb-FLAG (nsp1β), and analyzed by SDS-PAGE and autoradiography. Size markers and the position of bands for nsp2, nsp2TF, nsp2N and nsp1β are indicated next to each panel. *(B)* Band intensities were quantified by phosphor imaging and corrected for amino acid content and Met/Cys incorporation efficiency, after which the nsp2, nsp2TF and nsp2N levels were used to calculate ribosomal frameshifting efficiencies.

As before, PRRSV -2/-1 PRF was found to be highly efficient in this expression system. As seen in Fig. 2.7A, when only nsp2 was expressed, the non-frameshifted, full-length nsp2 constituted ~95% of the protein products immunoprecipitated with an antibody recognizing the N-terminal domain of nsp2. In the control expressing a self-cleaving nsp1β-nsp2 polyprotein from a single plasmid, there were equal amounts of nsp2 and nsp2TF produced and ~13% nsp2N. With WT nsp1β and nsp2 expressed from separate plasmids, ~56% was the -2 PRF product nsp2TF and ~9% was the -1 PRF product nsp2N. For three nsp1β mutants (Y131A, R134A and R142A), the level of -1/-2 PRF was as low as in the control expressing nsp2 only (Fig.2.7A) , highlighting the importance of these residues in PRF. The Y131A and R134A mutations in this nsp1β variant (from PRRSV isolate SD01-08) correspond to the Y125A and R128A mutations in nsp1β from PRRSV isolate SD95-21. For this previously used isolate, mutations Y125A and R128A were also found to almost completely abolish PRF stimulation (89, 180). For SD01-08 mutants K130A and R135A there was a significant reduction in both -2 (~70-80% reduction) and -1 PRF (~55-65% reduction). For Q137A, the reduction observed was less, with ~30% reduction in -2 PRF and a ~55% reduction for -1 PRF. The nsp1β expression level likely affects frameshifting efficiency, as previously described for viral protein 2A in the encephalomyocarditis virus. This protein acts as a PRF transactivator by binding to a genomic stem-loop structure, resulting in variable frameshift stimulation, from 0% at the start of infection to 70% late in infection as the concentration of 2A protein in the cells increased (121). Poor nsp1β expression could, therefore, result in reduced frameshift stimulation. The decrease in frameshifting that we observed for mutants K130A, Y131A, R134A, R135A, Q137A and R142A could, however, not be explained by insufficient expression of the nsp1β mutants in those samples because protein levels of these nsp1β mutants were comparable with WT nsp1β. The L132A and L136A mutants were expressed to lower levels

than WT, but for these mutants frameshift efficiencies were as high as with WT nsp1β so the amount of protein expressed was still sufficient for efficient frameshift stimulation (Fig. 2.7A).

2.3.8 Mutations in the nsp1β RBM motif affect innate immune suppression

Nsp1β and both -2/-1 PRF products, nsp2TF and nsp2N, have been implicated in suppressing host innate immune responses (131, 216, 219–221). Nsp1β may influence innate immune suppression in multiple ways. The protein was proposed to modulate the host immune response directly, but may also influence it indirectly through the -2/-1 PRF mechanism that directs nsp2TF and nsp2N expression. Specifically, the nsp1β RBM motif has been associated with both innate immune suppression and -2/-1 PRF stimulation (109, 122). When recombinant viruses with nsp1β RBM mutations are studied, it is not possible to establish whether phenotypic changes are caused by a reduced innate immune evasion capacity of nsp1β, altered -2/-1 PRF and nsp2TF/nsp2N expression levels, or a combination of the two. In order to study the different roles of nsp1β independently, it is important to uncouple its innate immune suppression function(s) from its PRF stimulatory activity. We have, therefore, also tested the impact of RBM mutations on nsp1β's ability to antagonize activation of the IFN-β response by using a dual luciferase reporter assay. Cells were co-transfected with plasmids expressing mitochondrial antiviral signaling protein (MAVS), which stimulates the pathway leading to IFN-β production, and either WT or mutant nsp1β. The inhibitory effect of nsp1β mutants on IFN-β promoter activation was measured via co-transfection of a firefly luciferase reporter gene construct under control of the IFN-β promoter. To correct for transfection efficiency variability, a plasmid encoding *Renilla* luciferase was co-transfected to provide an internal standard. At 18 h post-transfection, *Renilla* and firefly luciferase activities were measured. Activation of the IFN-β

promoter induced by MAVS expression only was set to 100%. As seen in Fig. 2.8, three mutants, K130A, Q133A and M141A, suppressed activation of the IFN-β promoter to an extent that was comparable to the suppression by WT nsp1β. Expression of mutants Y131A, R134A, R135A, V138A and G140A still allowed >50% of luciferase expression, indicating a strongly reduced ability to suppress IFN-β promoter activation. Mutants Y131A and R134A seem to be severely affected in both PRF stimulation (Fig. 2.7) and innate immune suppression (Fig. 2.8). Interestingly, mutant R142A, which was incapable of PRF stimulation, reduced luciferase expression by only 50%. Mutant K130A appeared to antagonize IFN-β activation even better than the wild type protein, while its reduction in -2 PRF stimulation is ~70%. For most other mutants some reduction in innate immune suppression capability was observed while PRF stimulation did not appear to be affected. Consequently, for future studies with recombinant viruses carrying nsp1ß mutations that reduce PRF stimulation, it may be advisable to use mutant K130A rather than Y131A, R134A or R135A, since the latter three mutations may also affect the protein's ability to counter innate immune responses in infected cells. The nsp1β mutants most able to suppress innate immune responses also suffer from strongly reduced -2/-1 PRF stimulation capability, which will complicate the assessment of the direct role of nsp1β in innate immune suppression during viral infection

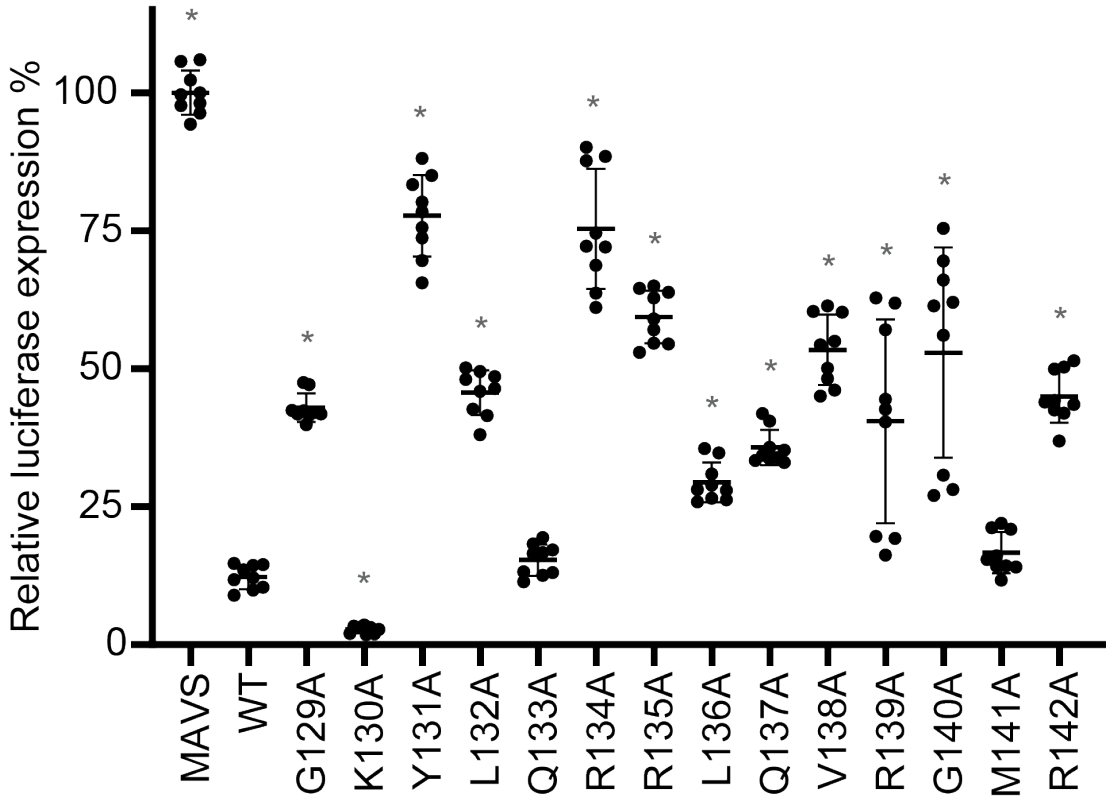


Figure 2.8 Analysis of innate immune supression by expression of PRRSV nsp1β mutants.

HEK-293T cells were co-transfected with plasmids expressing innate immune response inducer MAVS, a firefly luciferase reporter gene under control of the IFN-β promoter, *Renilla* luciferase and WT or mutant PRRSV nsp1β. Cells were lysed 18 h post transfection and the *Renilla and* firefly luciferase activities were measured. Firefly luciferase activity was normalized to *Renilla* activity in the same well. The average of three independent biological replicates with three technical replicates each is shown with the standard deviation. Significance ($p<0.001$) was assessed using an unpaired two-tailed Student's t test.

2.3.9 Structural insights into the -1/-2 PRF stimulatory complex

Having identified the minimal viral RNA sequence that forms a complex with nsp1β and PCBP2, we developed an approach to purify the protein-nucleic acid complex to assess its stoichiometry and structural biology. Assuming a 1:1:1 stoichiometry, nsp1β and PCBP2 were initially mixed in a 1:1 molar ratio at concentrations less that 1 mg/ml with a slight excess of nucleic acid (1.1 x molar excess) to generate the trimeric complex. After a 3-hour incubation period at 4 °C, the mixture was concentrated for loading on a gel filtration column. A fair amount of precipitation arose during this step, some of which may have been PCBP2 and nsp1β molecules that had not bound nucleic acid, as we found the proteins to be unstable in the absence of nucleic acid. Regardless, the resulting protein:RNA complex could be concentrated to 10 mg/ml at this stage, which was already an order of magnitude higher than the maximum concentrations of 1 mg/ml that could be achieved for nsp1β and PCBP2 on their own. The supernatant of the concentrated sample was separated from the precipitate and subsequently purified by size exclusion chromatography (Fig. 2.9A). SDS-PAGE and native PAGE were carried out on the purified samples to assess the composition of each complex, which revealed both proteins to be present (Fig. 2.9B), as well as the nucleic acid probe (Fig. 2.9C). Remarkably, the final purified complex could be concentrated to >20 mg/ml. The complex was found to be stable for at least 10 days at 4 °C.

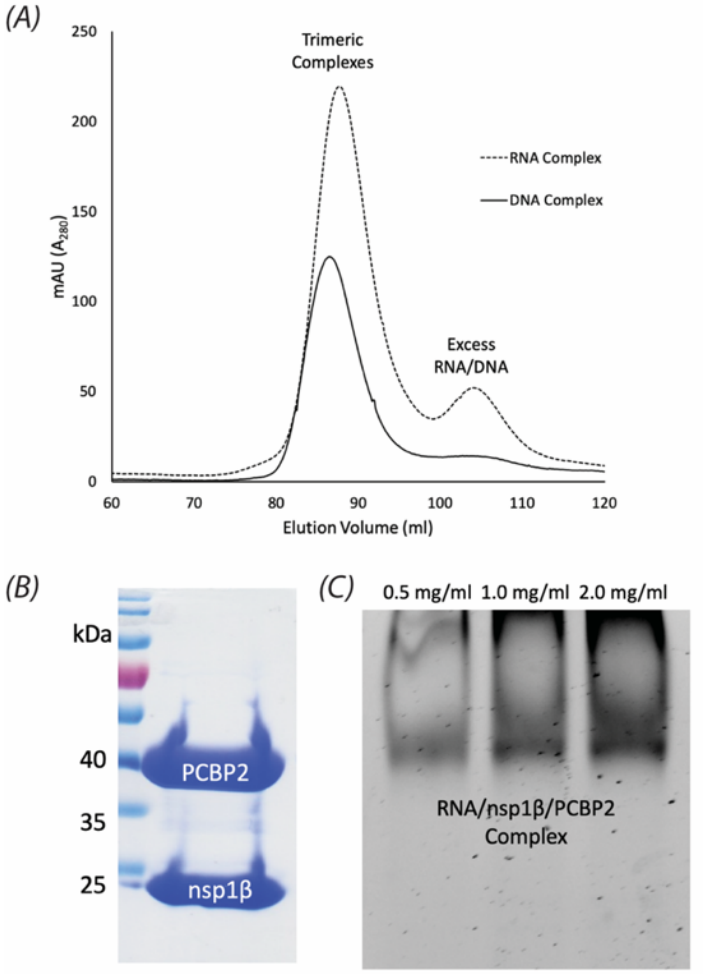


Figure 2.9 Purification of nsp1β:PCBP2:nucleic acid complexes for biophysical analysis.

Purified PRRSV nsp1β and PCBP2 were mixed together in equimolar concentrations with a 1.1-fold molar excess of the 34-nt ssRNA or ssDNA oligonucleotides that were identical or analogous, respectively, to the PRF signal in the nsp2-coding region of PRRSV SD01-08, as portrayed in Fig. 2.1B. Both quaternary complexes were found to be stable and could be concentrated to ~20 mg/ml prior to purification by size exclusion chromatography. *(A)* Elution trace of the protein:RNA complex from a Superdex200 gel filtration column monitored by UV light at 280 nm. *(B)* Coomassie Blue-stained SDS-PAGE of the purified complexes bound to RNA/DNA, revealing the presence of both proteins. *(C)* Non-denaturing 8% TBE PAGE gel of the purified complex shown in panel *(C)* stained with SYBR Gold, revealing the presence of the 34-nt RNA.

2.3.10 The Frameshift stimulatory complex exists in a 1:1:1 stoichiometry

To gain insight into the stoichiometry of the frameshifting complex, we characterized the nsp1β/PCBP2/ssRNA triple complex by the sedimentation velocity method using an analytical ultracentrifuge. The solvent and hydrodynamic parameters used during data analysis can be found in Appendix Table A1. We first measured a series of concentrations from 8 μM to 64 μM of the 34 nt ssRNA probe alone (Appendix Fig. A1 and Table A2). Two populations of species were apparent with roughly 80% of the material in species 1 and 20 % in species 2. This ratio remained constant over the concentration range that was investigated (Appendix Fig. A1 C). The experimental sedimentation constant s_e and hydrodynamic radius R_h of species 1 are nearly independent of concentration. Regarding species 2, s_e decreases very slightly and R_h decreases moderately with increasing concentration (Appendix Fig. A1 A-B). Concentration independence or a decrease of s_e and R_h is a sign that the kinetics of RNA chain exchange between the two species was very slow in relation to the time course of the experiment (24 h). Both species thus effectively acted like independent molecules. Self-interaction between macromolecules with a faster kinetics would manifest as an increase of s_e and R_h with increasing loading concentration.

If the partial specific volume $\bar{\nu}$ is known, the molecular mass M can be calculated from the extrapolated experimental sedimentation s^0_e and experimental diffusion constant D^0_e (the latter is represented here as hydrodynamic radius R_h^0) using the Svedberg equation. The extrapolation to infinite dilution is done to account for buffer effects and to eliminate effects from macromolecular self-association. For nucleic acids, however, $\bar{\nu}$ is a function of ionic strength and is expected to be in the range of 0.50 - 0.65 cm^3/g. Using a value of 0.628 cm^3/g we obtain a mass of 10.7 kDa for species 1, which corresponds to monomeric ssRNA. Dimeric or

higher order RNA could be ruled out as this would require a $\bar{v}$ outside the expected range. For species 2, this $\bar{v}$ value yielded a mass of 79.3 kDa and would correspond to an assembly of 7-8 RNA strands (Appendix Table A4). *c(s, f_r)* distributions, *c(s, M)* distributions and direct fitting of s_e and D_e using the Lamm equation (species analysis) are provided in Appendix Fig. A2-A5 together with the fit to the data and the residuals.

For the second step, we prepared three concentrations of 1.0, 2.0 and 4.0 mg/ml of the SEC-purified trimeric complex (Fig. 2.9A) and conducted sedimentation velocity experiments at 30,000 rpm. We also repeated the experiments at 42,000 rpm, using the previously used samples and cells. Comparing the total signal vs. loading concentration of both runs revealed a significant loss of signal in the second run (Appendix Fig. A6 A, D, G, and J). The increasing concentration of material at the bottom of the cell had led to irreversible aggregation and removal of material from the solution during the first run. We determined that the loading concentrations during the second run had reduced to 0.44, 0.75 and 1.09 mg/ml, thus expanding the investigated concentration range. Appendix Figs. A9, 10, 13, 14, 17 and 18, show the calculated 2-dimensional *c(s, f_r)* distributions obtained from the data with the sedimentation constant *s* on the x-axis and the diffusion constant expressed as frictional ratio f_r on the y-axis. A zoomed part of the plot was converted to mass and is represented as a two-dimensional *c(s,M)* distributions in the same figure. The one-dimensional distributions *c(s, *)*, *c(s)* and fit to the data together with residuals are shown as well. An overlay of the one-dimensional distributions of all concentrations are shown in Appendix Fig. A6. The *c(s, *)* distributions (Appendix Fig. A6 B, E, H and K) were obtained by integrating the 2-dimensional distributions along the f_r direction. Traditional *c(s)* distributions are also shown (Appendix Fig. A6 C, F, I and L), however these suffer from the incorrect assumption of an identical f_r value for all species. Both absorbance and interference optics of the XL-I

instrument were used, since the former is particularly sensitive to the nucleic acid due to its high extinction coefficient, whereas the interference optics is equally sensitive to all components. The $c(s, f_r)$ distributions obtained from the interference data in Appendix Fig. A6 E and K revealed three major, distinct populations of particles, especially in the data recorded during the first run. The absorbance optics could not resolve individual populations (Appendix. Fig. A6 B and H); however, comparing the one-dimensional distributions obtained from the absorbance optics to those obtained from the interference optics indicates that they cover the same *s* range. Thus, all three populations contained nucleic acid. Notably, the amount of nucleic acid decreases with increasing *s* values. Free RNA would show up at ~1.6 S. No such population was present, confirming that all RNA was bound to protein in the three populations of particles we observed.

To analyse the data in more detail, we directly fitted s_e and D_e of the observed populations (Fig. 2.10A & B, Appendix Fig. A7) using the Lamm equation. Depending on their sedimentation coefficients and hydrodynamic radii, we sorted the observed species into classes (Appendix Fig. A7 A and B). Three main species classes were present at all concentrations, with species class 1 and 2 contributing each ~ 30 - 40 % and species class 3 contributing ~ 10 - 20 % to the signal (Appendix Fig. A7 C). At loading concentrations from 0.754 mg/ml and higher we observed additional species classes with larger *s* values. Species class 4 contributed ~ 5 - 10 % to the signal. The remaining signal (5 - 15%) was shared by species classes 5 - 7 with very large sedimentation constants and very small hydrodynamic radii (equivalent to very large diffusion coefficients), either representing extreme shapes or non-ideality.

Surprisingly, *s* and R_h of species class 1, 2 and 3 remained constant or decreased with

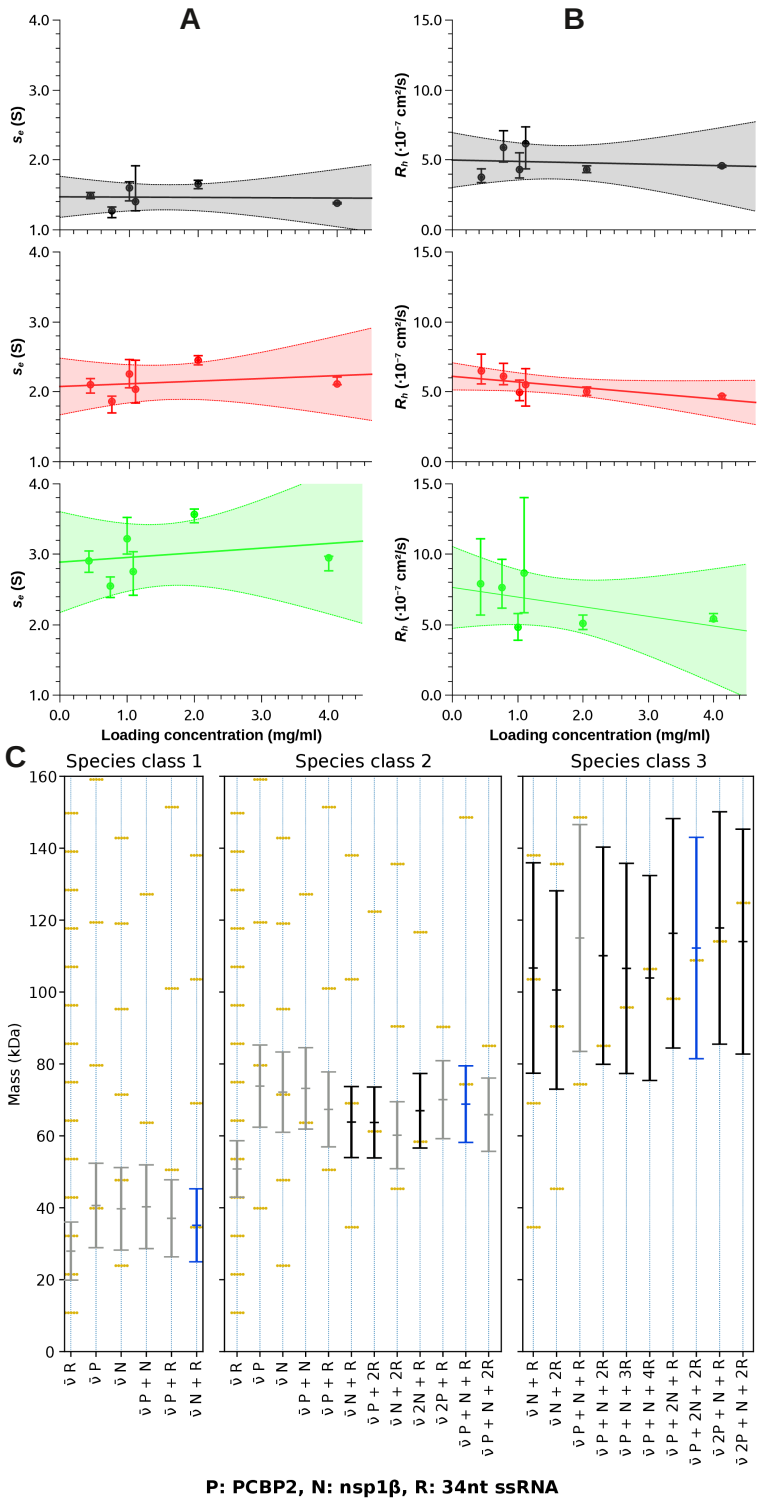


Figure 2.10 AUC of the trimeric complex shows a 1:1:1 stoichiometry.

(A, B) Results of the species analysis obtained by direct fitting the sedimentation coefficient *s* and diffusion coefficient *D* of each observed particle population to the Lamm equation (222) at each loading concentration. Only species classes 1 - 3 are shown. The panels show *(A)* experimental

sedimentation coefficient s_e and *(B)* D_e converted to hydrodynamic radius R_h versus loading concentration of species classes 1 (black), 2 (red) and 3 (green). Vertical error bars represent 95% confidence intervals of the fitted parameter (s_e or R_h). The values were then extrapolated to zero concentration using a unweighted linear fit (continuous line), yielding s^0_e and R_h^0. The shaded area shows the 95% confidence interval of the extrapolation. *(C)* Conversion of s^0_e and D^0_e to mass. The conversion relies on the partial specific volume $\bar{v}$ which depends on the ratio of the components of the complex which is shown at the bottom of the plot. Multiples of the same ratio have the same $\bar{v}$; the corresponding mass ladders are shown as golden rungs. Due to the ambiguity of $\bar{v}$, multiple solutions are possible. From the absorbance optics we know that species 1 - 3 must contain RNA. All solutions without RNA and solutions that do not intersect with a mass ladder rung can be excluded (grey bars). We have marked the solution for each species class we deem the most likely with a blue bar. Black bars show alternate possible solutions. The error bars represent 95% confidence intervals which are based on the experimental uncertainties of s^0_e, D^0_e and the solvent density and viscosity.

increasing loading concentration (Fig. 2.10 A & B), indicating stable particles with a very slow exchange of components with the other species (relative to the time course of the experiment). We could therefore extrapolate the values to infinite dilution and determine s^0_e, D^0_e and $R_h{}^0$ (Fig. 2.10A & B, and Appendix Table A3, which shows the values converted to standard conditions).

As mentioned earlier, s^0_e and $D^0{}_e$ can be converted to mass if the $\bar{\nu}$ is known. nsp1β and PCBP2 were produced by bacterial expression, therefore not glycosylated and their $\bar{\nu}$ could be accurately predicted from the amino acid sequence (197). As described above, we measured the $\bar{\nu}$ of the 34nt ssRNA probe alone in the same buffer environment in which we produced the complex. We thus knew the $\bar{\nu}$ of every component and could calculate the resulting $\bar{\nu}$ and mass of each species class for every conceivable composition. Fig. 2.10C shows a collection of conceivable masses for species classes 1 – 3.

Particles in species class 1 were roughly half the molecular weight of those in species class 2. As the absorbance optics tell us, there must be nucleic acid and protein present within all three species. As seen in Fig. 2.10C (species class 1), the only possible composition involving nucleic acid could be a monomer of nsp1β bound to a monomer of RNA (total mass of ~35 kDa). This was unexpected since neither protein alone appears to bind RNA in EMSA assays (Fig. 2.5). Nevertheless, there is precedence for this behaviour due to the presence of the RBM of nsp1β and it is possible this interaction is not detectable by EMSA. This suggests that nsp1β directly interacts with nucleic acid in the final tricomponent system.

Regarding species class 2, more than a single stoichiometry of the components would match the experimentally determined mass (*e.g.* nsp1β:RNA 2:2 or PCBP2:RNA 1:2). However, the eluate from the size exclusion column contains both, nsp1β and PCBP2 (Fig. 2.9) and PCBP2 on its own is not competent to bind the RNA probe. We, therefore, deem that the

nsp1β:PCBP2:RNA 1:1:1 stoichiometry (triple complex) that also falls within the 95 confidence interval is the most likely solution.

The third species class was present in lower abundance (10 - 20 %), which resulted in a noisier signal and, therefore, a larger uncertainty of the mass, which falls in the range from 75 - 150 kDa, depending on $\bar{\nu}$ and the size of the confidence interval (Fig. 2.10C). Still, the mass is outside of the confidence interval of a super-complex composed of two copies of the tricomponent complex in a 1:1:1 stoichiometry (~150 kDa). We theorize that two independent monomers of nsp1β have bound two individual monomers of the RNA probe and they are held together by one copy of PCBP2 to give a mass of ~ 110 kDa. It is possible that PCBP2 is tethering together this super-complex by utilizing both the KH1 and KH3 domains independently; however, given its low abundance and stoichiometry, this complex may not be biologically relevant.

The above results suggest that nsp1β from PRRSV isolate SD01-08 from species PRRSV-1 exists as a monomer both in solution and in the PRF stimulatory complex; however, nsp1β from PRRSV isolate XH-GD from species PRRSV-2 is reported as a dimer according to its X-ray structure (PDB ID: 3MTV(92)). The two proteins share an amino acid sequence identity of 40 % (EMBOSS Needle (223)), and of the 32 amino acids identified by PISA (224) in the proposed homodimer interface of XH-GD nsp1β, 12 (38%) are conserved with nsp1β from SD01-08. Given our observations, the multimeric state of nsp1β may differ between the two isolates; however, it is also possible that nsp1β from PRRSV isolate XH-GD dimerizes at concentrations required for its crystallization.

2.3.11 Small-angle X-ray scattering supports a 1:1:1 binding stoichiometry for the nsp1β:PCBP2:ssRNA species 2 complex

To further understand the structure of the nsp1β:PCBP2:RNA complex, a three-dimensional molecular envelope of the complex was determined experimentally by SAXS, which provides insights into the low-resolution structural information of biomolecules and their complexes under physiological buffer conditions. We used a HPLC-SAXS setup to collect scattering data for the complex of nsp1β:PCBP2:ssRNA. The X-ray scattering trace and UV traces of the 10 mg/ml sample eluting from the 4.5 ml Shodex KW40 column are shown in Appendix Fig. A21. Consistent with the sedimentation velocity results, the UV trace reveals some heterogeneity of the sample. However, the main peak of the X-ray trace is symmetric and homogeneous and originates predominantly from a single species. The buffer subtracted and merged SEC-SAXS data taken from frames at the peak center are presented in Fig. 2.11A. Next, we performed the Guinier analysis of merged data to ensure the purity of the complex and to determine the R_g (radius of gyration) from the SAXS data belonging to the low-q region (225). Inset to Fig. 2.11A represents the Guinier plot for the complex with a linear region at low-q values indicating that the complex is monodispersed. The Guinier analysis for the complex also provided an R_g value of 3.900 ± 0.011 Å (Table 2.4). Next, we processed the SAXS scattering data from Fig. 2.11A to perform Kratky analysis to investigate the folding state of biomolecules (226, 227). The globular-shaped biomolecules typically display a well-defined maximum value of 1.1 at q^*R_g = 1.73 (228). As presented in Fig. 2.11B, the Kratky analysis for the complex under investigation suggest that it is well folded and has extended conformation in solution. Now that we confirmed homogeneity and folded state of this complex, we converted the SAXS raw data into the real space

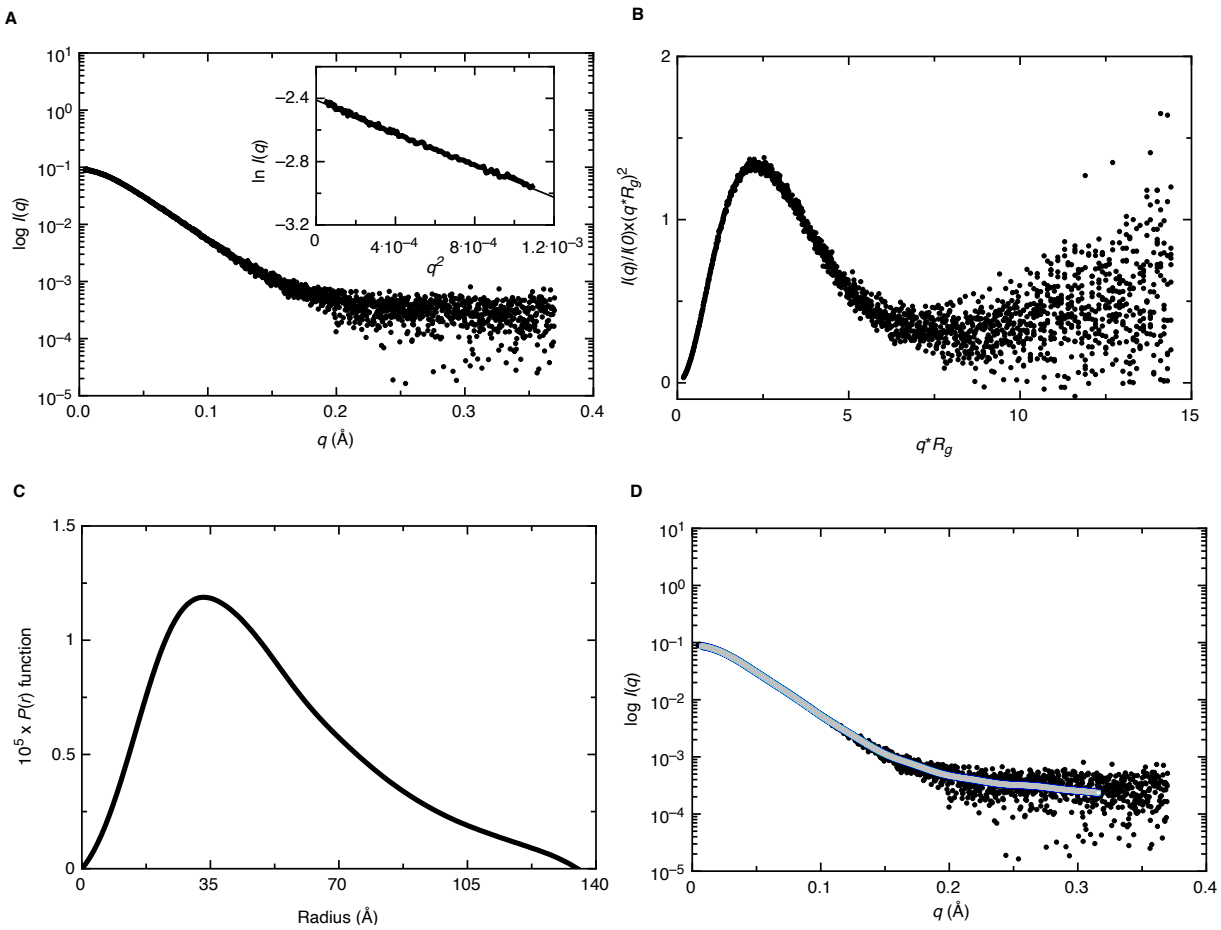


Figure 2.11 Characterization of the PRRSV nsp1β:PCBP2:RNA complex using SAXS.

(A) A plot of scattering intensity versus scattering angle presenting the merged data for nsp1β:PCBP2:ssRNA. Inset to this plot is the Guinier analysis which confirms the homogeneity of the complex. *(B)* Dimensionless Kratky plot for the nsp1β:PCBP2:ssRNA demonstrating its extended structure in solution. *(C)* A pair-distance distribution ($P(r)$) plot for nsp1β:PCBP2:ssRNA complex allowing the determination of R_g and D_{max}. *(D)* Alignment between experimentally collected SAXS data (dark circles) and calculated data from representative low-resolution models (solid lines).

Table 2.4 Biophysical parameters of the PRRSV nsp1□:PCBP2:ssRNA 1:1:1 complex

Parameter	*Experimental value*	*Experimental method*	*Dammin Models*
Sedimentation coefficient $s^0_{20°C,\,w}$ (S)	2.84 [2.28 - 3.41]	SV	4.22 ± 0.02††
Hydrodynamic radius $R_h{}^0_{20°C,\,w}$ (nm)	6.11 [5.17 - 7.46]	SV	4.40 ± 0.02††
Molecular mass *M*† (kDa)	68.8 [58.2 - 79.5]	SV	
Molecular mass *M* (kDa)	86 ± 6	SAXS - Primus	
Formula mass (Da)	*74224.49*	-	
Extrapolated scattering intensity at 0 angle *I(0)*	(0.089 ± 1.60) $\cdot 10^4$	SAXS - Guinier	
Radius of gyration R_g (nm)	3.900 ± 0.011	SAXS - Guinier	
Extrapolated scattering intensity at 0 angle *I(0)*	(0.089 ± 1.35) $\cdot 10^4$	SAXS - *P(r)*	
Radius of gyration R_g (nm)	4.002 ± 0.008	SAXS - *P(r)*	4.044 ± 0.003
Longest dimension D_{max} (nm)	13.5	SAXS - *P(r)*	13.98 ± 0.02
Volume *V* (nm^3)		SAXS	165 ± 1
χ2 of fit		SAXS	~1.3
Normalised spatial discrepancy *NSD*		SAXS	0.60 ± 0.01

†using $\bar{\nu} = 0.7166\ cm^3/g$

††calculated in Hydropro

Uncertainties are given as 95% confidence intervals.

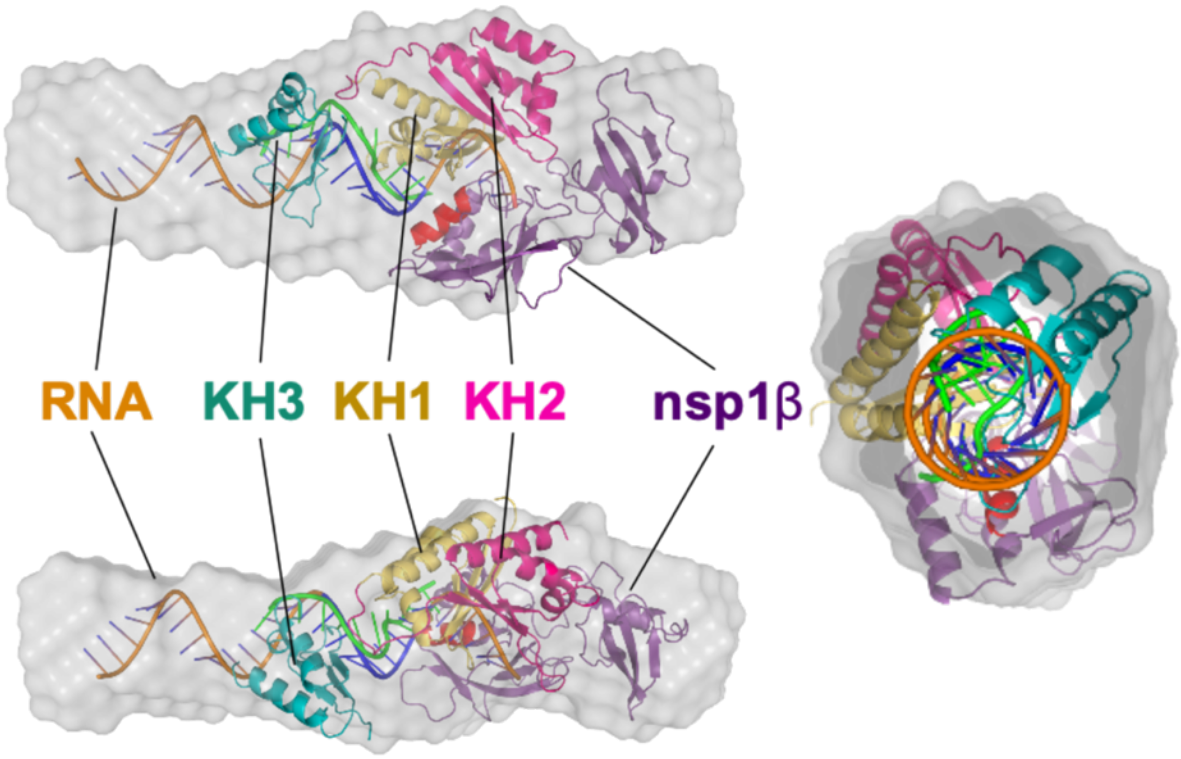


Figure 2.12 Low resolution SAXS structure of the trimeric complex is consistent with a 1:1:1 stoichiometry.

The SAXS envelope is shown in three different orientations with the calculated density filled in with existing protein structures and modeled nucleic acid. In purple is a monomer of nsp1β from PRRSV strain XH-GD (PDB ID: 3MTV (183)) with the theorized RBM helix depicted in red. In gold and fuchsia are the KH1 and KH2 domains of PCBP2, respectively (PDB ID: 2JZX (206)). In green is shown a short, C-rich RNA motif that was co-crystalized with KH1 alone (PDB ID: 2PY9 (207)), which is shown as a frame of reference for how our idealized RNA probe (orange) may fit. In teal is the KH3 domain of PCBP2 (PDB ID: 2P2R (210)). The RNA molecule is shown in orange with the C-rich motif shown in blue. Figures were generated with PyMOL (209).

electron pair-distance distribution function ($P(r)$ as presented in Fig. 2.11C using the program GNOM (203). We also obtained R_g and D_{max} (maximum particle dimension) values of 4.002 ± 0.008 nm and 13.5 nm respectively (Table 2.4). The R_g from Guinier analysis which only takes in account the low-q region and from $P(r)$ analysis that utilizes a wider range of SAXS data agree with each other, as well as the Kratky analysis suggests that the complex is folded, implying that we have suitable data for low-resolution structure analysis. We utilized the DAMMIN program (204) to calculate 20 models for the nsp1β:PCBP2:nucleic acid complex, as described earlier (229, 230). The χ^2 values in each case were ~1.3 representing a good agreement between the experimentally collected and low-resolution structure-derived data (Table 2.4), as presented in Fig. 2.11D. Finally, we used the DAMAVER package (205) to rotate and align all 20 low-resolution structures and to obtain an averaged filtered structure of the complex. The normalized spatial discrepancy (NSD) parameter in DAMAVER describes the goodness of the superimposition of individual models. For the nsp1β:PCBP2:nucleic acid complex we obtained an NSD value of 0.60 ± 0.01 indicating that all 20 low-resolution structures are very similar to each other. DAMMIN reported an R_g of 4.044 ± 0.003 nm and a D_{max} of 13.98 ± 0.02 nm for its models, which agrees with the numbers obtained from GNOM analysis (Table 1). We also performed HYDROPRO (231) calculations of all 20 low-resolution structures using the partial specific volume of the triple complex to calculate the sedimentation coefficients and hydrodynamic radii of the models for comparison with the experimental data (Table 2.4) (232).

Fig. 2.12 presents the elongated averaged filtered structures for the nsp1β:PCBP2:nucleic acid complex into which existing 3D structures of nsp1β (PDB ID: 3MTV(183)), the KH1-KH2 region (PDB ID: 2JZX) (206) and KH3 (PDB ID: 2P2R) of PCBP2, respectively, were fitted along with a model of the 34-nt ssRNA probe (generated by w3DNA 2.0 (208)). Given our sedimentation

velocity data above, the crystal structure of nsp1β (from PRRSV strain XH-GD(183)) was fitted as a monomer. The NMR solution structure of the PCBP2 KH1-KH2 fusion was also fitted along with the X-ray structure of the third KH domain (PDB ID: 2P2R) (210). It was previously found that KH1 and KH3 participate in nucleic acid binding, whereas KH2 does not (123). Given this constraint, the nucleic acid binding regions of KH1 and KH3 were oriented toward the C-rich region of the ssRNA within the modeled complex, as was the RBM helix of nsp1β. The relative position and orientation of KH1, KH2 and KH3 of PCBP2 to each other and to the ssRNA probe were modeled based on the X-ray structure of KH1 bound to RNA (PDB ID: 2PY9 (207)). All fitting was carried out in PyMOL (209). As shown in Fig. 2.12, the SAXS envelope is tubular with a distinct bulge at one end. An idealized helical model of the 34 nt ssRNA fits within the tubular portion of the envelop, while the bulge is large enough to account for one molecule each of nsp1β and PCBP2, which is in agreement with our AUC findings of a 1:1:1 stoichiometry. It should be noted that in the X-ray structure of KH1 bound to RNA, the RNA does not adopt a perfectly helical conformation and is instead more linear (207). Thus, it is likely that the RNA within the nsp1β:PCBP2:RNA complex is also not perfectly helical throughout. Nevertheless, the resolution of our SAXS data is insufficient to gain insight into the true conformation of the ssRNA probe and thus was left in an ideal helical conformation. Further, 3D structural information is not available for residues 170-287 of PCBP2 that span between the KH2 and KH3 domains. This 117 amino acid region accounts for ~ 30 % of the total PCBP2 structure and is thought to be highly disordered (206, 210, 211). In keeping with this, the SAXS envelope does not appear to account for this mass, which we assume is due to the high degree of disorder in this region of PCBP2.

Finally, the C-rich motif, and presumed binding site of the nsp1β:PCBP2 complex, is positioned near the 3'-end of the 34-nt ssRNA probe. This position is consistent with the location

of the bulge that appears in the SAXS envelope and thus where we believe nsp1β and PCPB2 bind to the RNA (Fig 2.1B). The co-localization of nsp1β and PCPB2 in the model is consistent with our hypothesis that both nsp1β and PCBP2 interact directly with the viral RNA genome at the C-rich motif.

2.4 Conclusion

Together, our results provide new structural and functional insights into the unique PRF mechanism that is employed by arteriviruses, in which a viral and a host protein cooperate with a specific signal in the viral RNA genome to direct the expression of two additional viral protein species. A number of residues within both nsp1β and PCBP2 are required for a nucleic acid binding event that triggers the frameshifting during PRRSV genome translation. We believe that this complex is dynamic and must be able to readily assemble and disassemble in order to interact with the ribosome and thereby facilitate ribosomal stalling, which allows for PRF and subsequently detach from the genome to allow for downstream translation. Sedimentation analysis by analytical ultracentrifugation revealed that nsp1β and PCBP2 each bind to the viral RNA genome as monomers, which is consistent with our structural analysis for the complex by SAXS. Interestingly, the monomeric form of nsp1β in the PRF complex differs from a previously determined X-ray structure of nsp1β, which appears as a dimer. The ability of nsp1β to form a dimer versus its monomeric interaction with PCBP2 may underlie a mechanism that regulates the frequency of PRF during virus replication. Future X-ray crystallographic and cryo-EM studies will hopefully reveal the finer structural details of this fascinating example of the non-canonical translation of viral mRNAs.

2.5 Future directions

It has been shown that not only PCBP2 acts as a transactivator of PRF, but also PCBP1 as well (123). These proteins are ~83 % identical at the amino acid level. It would be prudent to carry out similar mutational analysis as was done with PCBP2. R40, R57 and N325 are all conserved amongst the two proteins and would likely show similar results. The lone crystal structure of PCBP1 is that of its first KH domain bound to C-rich nucleic acid (PDB: 1ZTG) (214). The structure revealed that both R40 and R57 both participate in nucleic acid binding, which was the case for PCBP2 binding our nucleic acid probes; however, three alternate residues (Gly22, Ile49 and Asp82) also form hydrogen bonds with cytosine bases (214). All three of these residues are conserved within PCBP2. Carrying out these additional mutations in both PCBP1 and 2 to assess complex formation via EMSAs would be a sensible next step to better understand which residues of the PCBPs can stimulate PRF within the PRRSV genome.

Although high-resolution structural analysis of the tricomponent complex through protein crystallography would be ideal, the surge in recent advancements in cryo-electron microscopy (Cryo-EM) has made it a very powerful tool for structural characterization of macromolecules (233). Although this complex (~ 72 kDa) may be on the edge for detection limit by the current standards of cryo-EM, it is absolutely worth attempting. Alternatively, the size of the ribosome can be exploited to increase the molecular mass of the complex significantly for easier detection of the tricomponent complex, as this complex would likely be recognized by the ribosome. A recent study showed nsp1 of SARS-CoV-2 (~20 kDa) bound to the ribosome, as it acts to redirect translational machinery of infected host cells towards viral mRNA. The

structure was determined down to 2.5 Å resolution (234). Protein cross-linking could also be attempted as this may be a rather dynamic system and would help lock the complex into place.

2.6 Appendix

Table A.1

Parameters	
Solvent **1x PBS (pH 7.4 @ 25°C), 100 mM KCl, 5% glycerol**	2.027 mM sodium dihdyrogen phosphate 7.973 mM disodium hydrogen phosphate 138 mM sodium chloride 102.7 mM potassium chloride ~684 mM glycerol
Solvent density[1] ρ (g/cm^3)	1.0354 ± 0.0005
Buffer viscosity[1] η (P)	0.01235 ± 0.0002
Temperature T (°C)	20.0
his$_6$-PCBP2 partial specific volume[1] $\bar{\nu}$ (cm^3/g)	0.73318
his$_6$-PCBP2 formula mass[2] M_w (kDa)	39749.32
his$_6$- PCBP2 absorption coefficient[2] ε_{280} (M^{-1} cm^{-1})	14815
PRRSV nsp1β partial specific volume[1] $\bar{\nu}$ (cm^3/g)	0.72777
PRRSV nsp1β formula mass[2] M_w (kDa)	23786.87
PRRSV nsp1β absorption coefficient[2] ε_{280} (M^{-1} cm^{-1})	46200
34nt ssRNA partial specific volume $\bar{\nu}$ (cm^3/g)	dependent on cation concentration expected range 0.50 - 0.65
34nt ssRNA formula mass[3] M_w (kDa)	10688.3
34nt ssRNA absorption coefficient[3] ε_{260} (M^{-1} cm^{-1})	323000

[1]Sednterp, [2]ProtParam, [3]IDTDNA

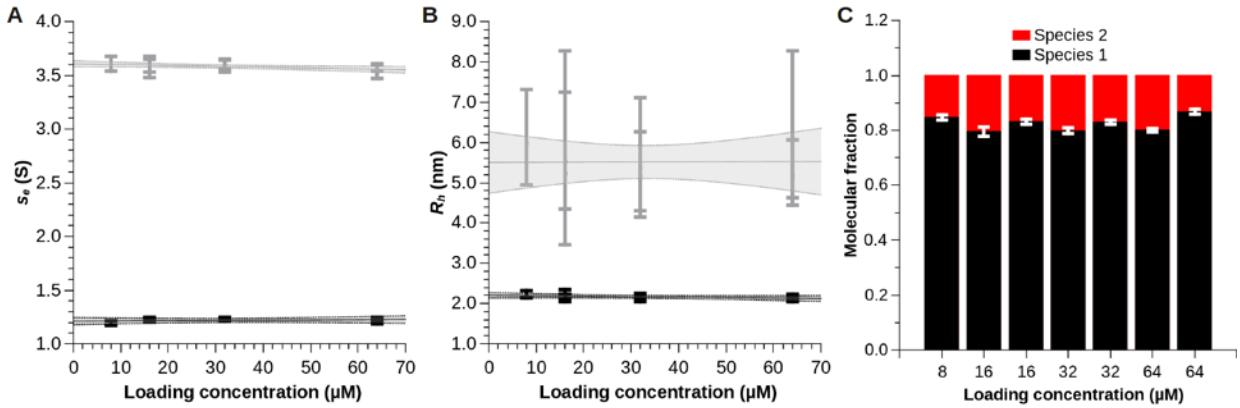


Fig. A1

Table. A2

Results

Species 1	
Sedimentation coefficient s^0_e	1.21 ± 0.04 S
Diffusion coefficient D^0_e	7.9 ± 0.3 ·10⁻⁷ cm²/s
Partial specific volume that yields correct mass (10.7 kDa)	0.62763 [0.60703 - 0.64704] cm^3/g
Sedimentation coefficient $s^0_{20°C,\,w}$	1.59 [1.54 - 1.65] S
Hydrodynamic radius R_h^0	2.20 [2.12 - 2.29] nm
Species 2	
Sedimentation coefficient s^0_e	3.61 ± 0.03 S
Diffusion coefficient D^0_e	3.2 ± 0.5 ·10⁻⁷ cm²/s
Sedimentation coefficient $s^0_{20°C,\,w}$	4.74 [4.65 - 4.83] S
Hydrodynamic radius R_h^0	5.49 [4.84 - 6.35] nm
Molecular mass if RNA; would contain 7-8 strands	79.3 [68.2 - 90.5] kDa

Uncertainties are given as 95% confidence intervals

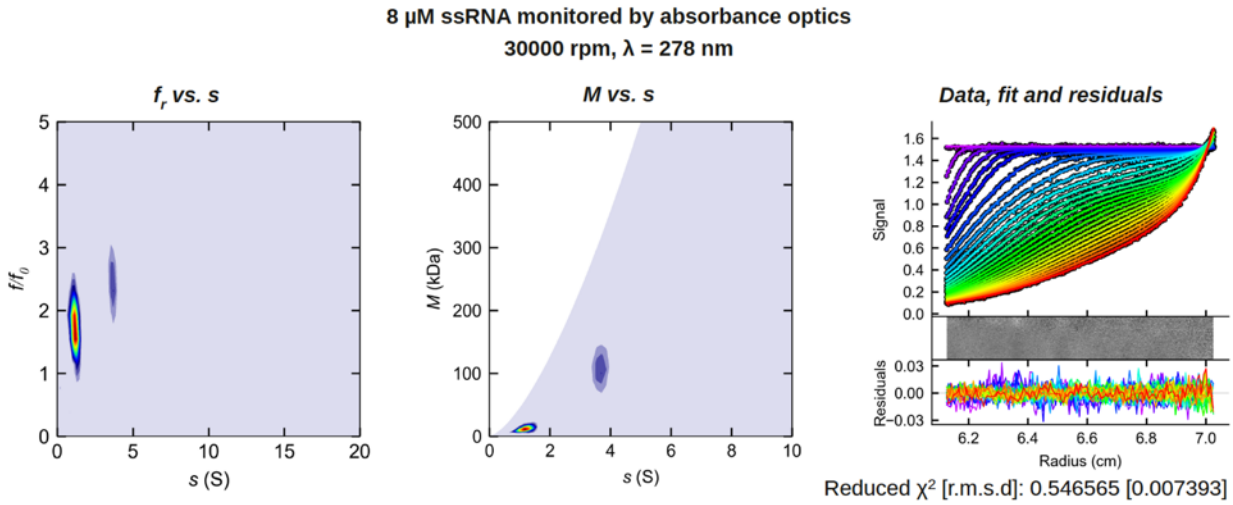
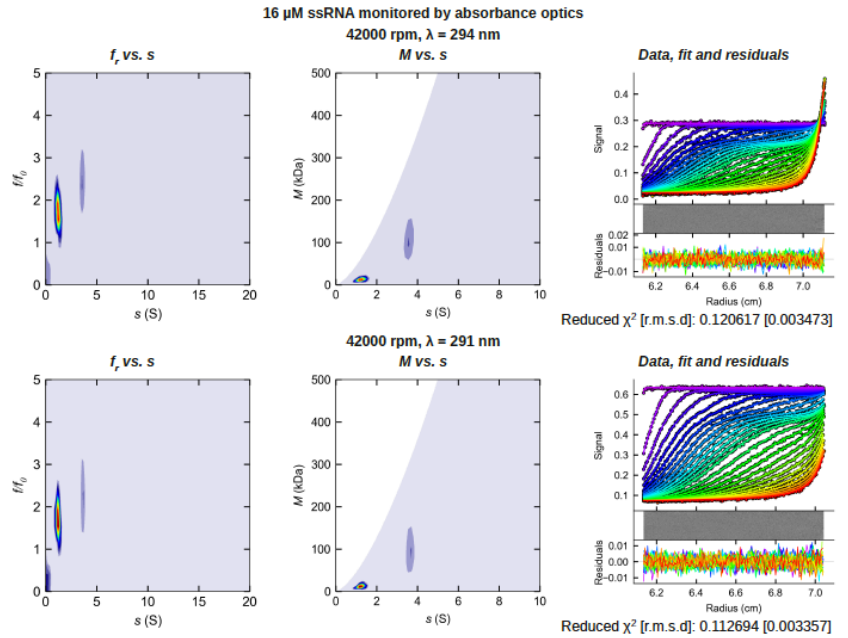
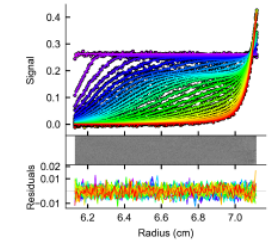

8 µM ssRNA monitored by absorbance optics
30000 rpm, λ = 278 nm
f_r vs. s
f/f_0
s (S)
M vs. s
M (kDa)
s (S)
Data, fit and residuals
Signal
Residuals
Radius (cm)
Reduced χ² [r.m.s.d]: 0.546565 [0.007393]

Fig. A2



Species Analysis (42000 rpm, λ = 294 nm, exp. # 1)

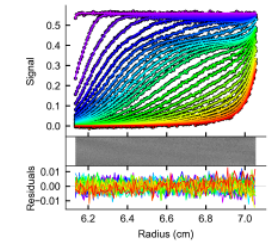
Parameter	Value	95% confidence interval
Species 1		
Sedimentation coefficient s_w (S)	1.229725	1.2078 - 1.2518
Diffusion coefficient D_o ($\cdot 10^{-7}$ cm²/s)	7.837109	7.4467 - 8.2514
Species 2		
Sedimentation coefficient s_w (S)	3.579254	3.4808 - 3.6784
Diffusion coefficient D_o ($\cdot 10^{-7}$ cm²/s)	3.316965	2.1005 - 5.0301
Fraction (signal)	0.204758	0.1884 - 0.2216
Common		
Total concentration (signal)	0.2623926	0.2545 - 0.2705
Meniscus position (cm)	6.113078	6.1073 - 6.1188
Bottom position (cm)	7.227485	7.2230 - 7.2323
Reduced χ2 of fit [rmsd]	0.1228484	
R.m.s.d. of fit	0.003504974	



Reduced χ² [r.m.s.d]: 0.122848 [0.003505]

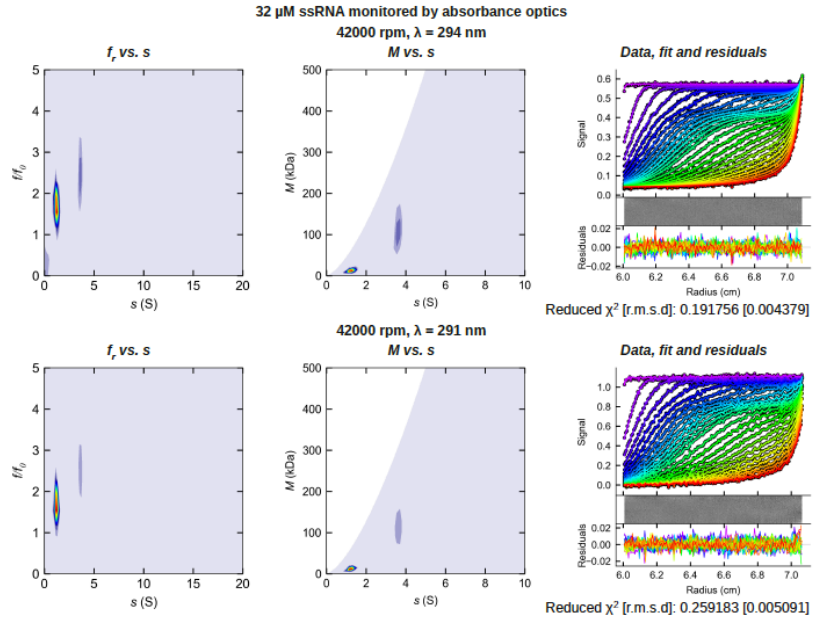
Species Analysis (42000 rpm, λ = 291 nm, exp. # 2)

Parameter	Value	95% confidence interval
Species 1		
Sedimentation coefficient s_w (S)	1.215424	1.2044 - 1.2265
Diffusion coefficient D_o ($\cdot 10^{-7}$ cm²/s)	8.173451	7.8048 - 8.4944
Species 2		
Sedimentation coefficient s_w (S)	3.590108	3.5299 - 3.6475
Diffusion coefficient D_o ($\cdot 10^{-7}$ cm²/s)	3.140034	2.3970 - 4.0048
Fraction (signal)	0.168034	0.1588 - 0.1769
Common		
Total concentration (signal)	0.5605898	0.1588 - 0.1769
Meniscus position (cm)	6.117462	6.1147 - 6.1205
Bottom position (cm)	7.234521	7.2271 - 7.2411
Reduced χ2 of fit [rmsd]	0.1260963	
R.m.s.d. of fit	0.003551004	



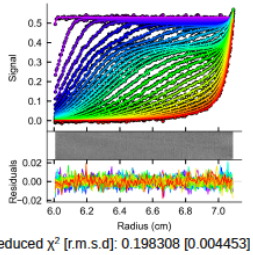
Reduced χ² [r.m.s.d]: 0.126096 [0.003551]

Fig. A3



Species Analysis (42000 rpm, λ = 294 nm, exp. # 1)

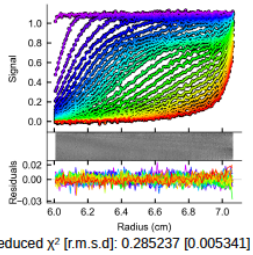
Parameter	Value	95% confidence interval
Species 1		
Sedimentation coefficient s_w (S)	1.232830	1.2201 - 1.2468
Diffusion coefficient D_s (·10^{-7} cm²/s)	8.121437	7.7671 - 8.5080
Species 2		
Sedimentation coefficient s_w (S)	3.585519	3.5308 - 3.6467
Diffusion coefficient D_s (·10^{-7} cm²/s)	3.231844	2.4451 - 4.1974
Fraction (signal)	0.202133	0.1919 - 0.2126
Common		
Total concentration (signal)	0.5306498	0.5205 - 0.5417
Meniscus position (cm)	5.986057	5.9823 - 5.9896
Bottom position (cm)	7.236567	7.2309 - 7.2428
Reduced χ2 of fit [rmsd]	0.1983076	
R.m.s.d. of fit	0.004453174	



Reduced χ² [r.m.s.d]: 0.198308 [0.004453]

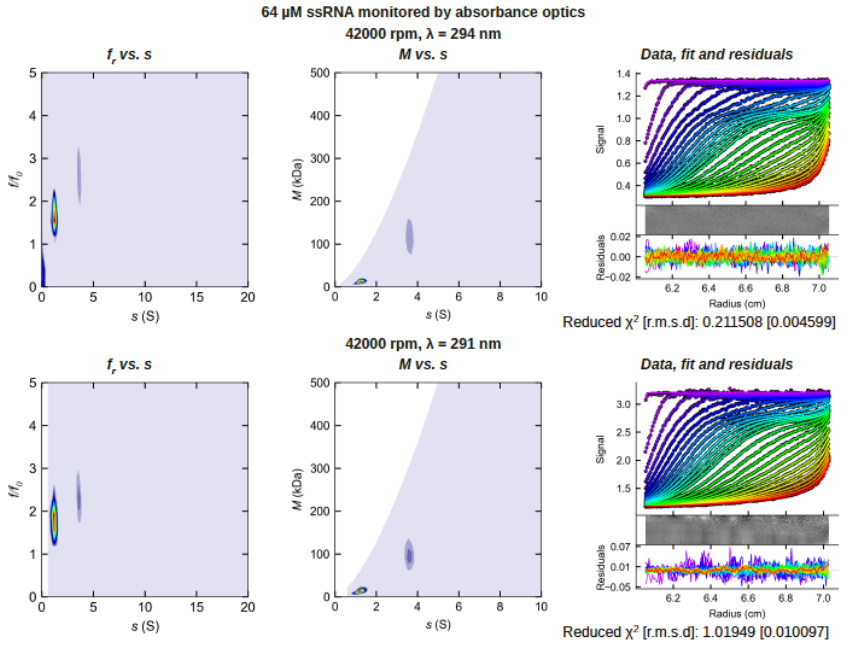
Species Analysis (42000 rpm, λ = 291 nm, exp. # 2)

Parameter	Value	95% confidence interval
Species 1		
Sedimentation coefficient s_w (S)	1.220184	1.2127 - 1.2280
Diffusion coefficient D_s (·10^{-7} cm²/s)	7.938753	7.7172 - 8.1994
Species 2		
Sedimentation coefficient s_w (S)	3.596446	3.5560 - 3.6386
Diffusion coefficient D_s (·10^{-7} cm²/s)	3.366156	2.7780 - 4.0454
Fraction (signal)	0.170676	0.1640 - 0.1770
Common		
Total concentration (signal)	1.088870	1.0756 - 1.1033
Meniscus position (cm)	5.989551	5.9874 - 5.9917
Bottom position (cm)	7.235867	7.2314 - 7.2412
Reduced χ2 of fit [rmsd]	0.2852372	
R.m.s.d. of fit	0.005340760	



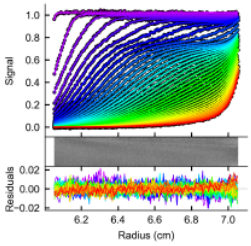
Reduced χ² [r.m.s.d]: 0.285237 [0.005341]

Fig. A4



Species Analysis (42000 rpm, λ = 294 nm, exp. # 1)

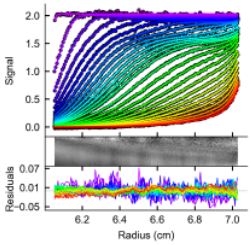
Parameter	Value	95% confidence interval
Species 1		
Sedimentation coefficient s_w (S)	1.239898	1.2324 - 1.2474
Diffusion coefficient D_e (·10^{-7} cm²/s)	8.170592	7.9399 - 8.4006
Species 2		
Sedimentation coefficient s_w (S)	3.572412	3.5362 - 3.6070
Diffusion coefficient D_e (·10^{-7} cm²/s)	3.355237	2.8711 - 3.9210
Fraction (signal)	0.198935	0.1930 - 0.2052
Common		
Total concentration (signal)	1.024325	1.0124 - 1.0365
Meniscus position (cm)	6.025673	6.0235 - 6.0279
Bottom position (cm)	7.241323	7.2362 - 7.2464
Reduced χ2 of fit [rmsd]	0.2163596	
R.m.s.d. of fit	0.004651447	



Reduced χ² [r.m.s.d]: 0.21630 [0.004651]

Species Analysis (42000 rpm, λ 291 nm, exp. # 2)

Parameter	Value	95% confidence interval
Species 1		
Sedimentation coefficient s_w (S)	1.201514	1.1918 - 1.2113
Diffusion coefficient D_e (·10^{-7} cm²/s)	8.107909	7.7838 - 8.4409
Species 2		
Sedimentation coefficient s_w (S)	3.534074	3.4728 - 3.5943
Diffusion coefficient D_e (·10^{-7} cm²/s)	2.847448	2.1050 - 3.7689
Fraction (signal)	0.132244	0.1238 - 0.1408
Common		
Total concentration (signal)	2.030704	1.9974 - 2.0654
Meniscus position (cm)	6.027494	6.0247 - 6.0302
Bottom position (cm)	7.245158	7.2363 - 7.2544
Reduced χ2 of fit [rmsd]	1.393352	
R.m.s.d. of fit	0.011804033	



Reduced χ² [r.m.s.d]: 1.393352 [0.011804]

Fig. A5

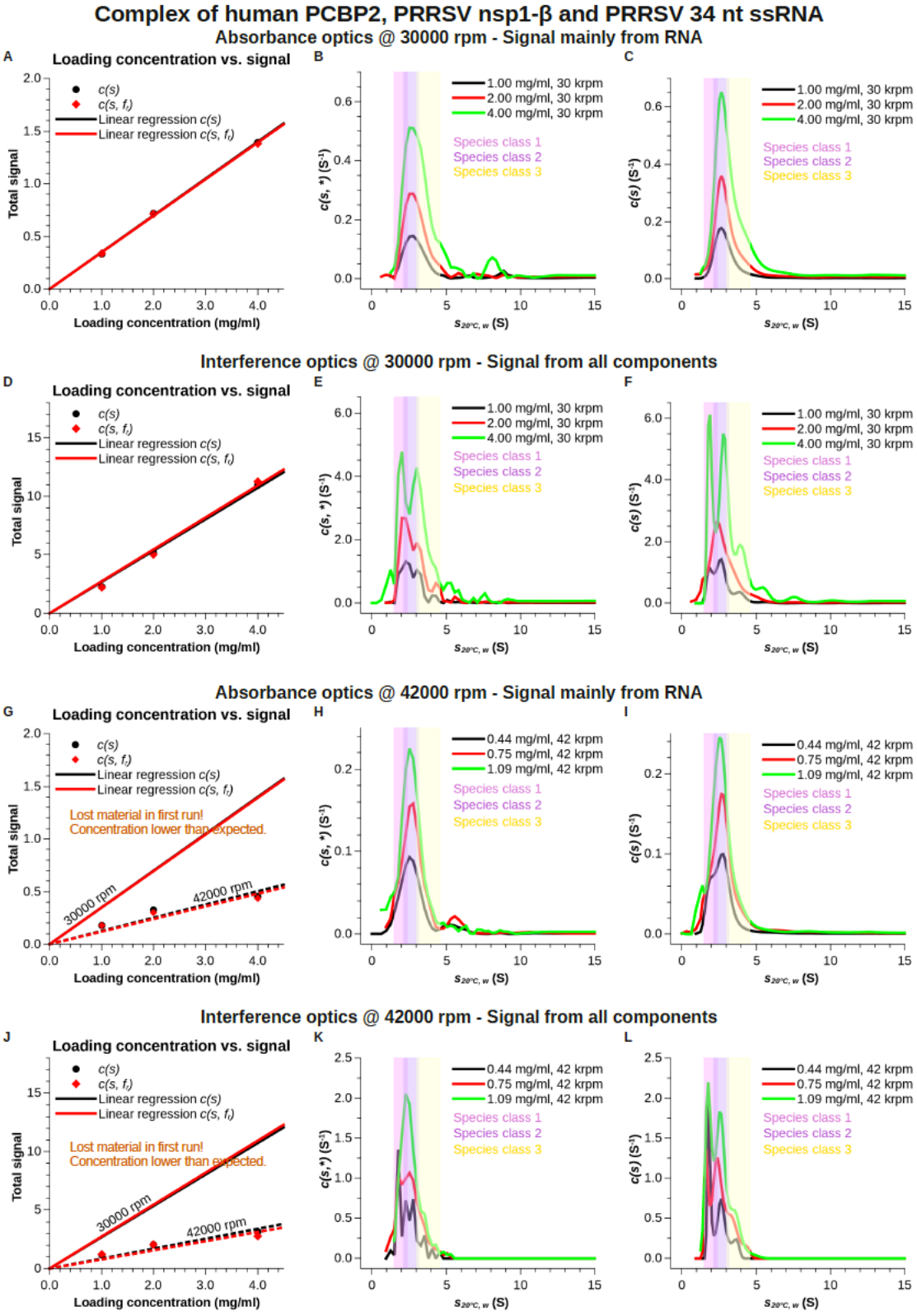

Complex of human PCBP2, PRRSV nsp1-β and PRRSV 34 nt ssRNA
Absorbance optics @ 30000 rpm - Signal mainly from RNA
A
Loading concentration vs. signal
c(s)
c(s, f_r)
Linear regression c(s)
Linear regression c(s, f_r)
Total signal
Loading concentration (mg/ml)
B
1.00 mg/ml, 30 krpm
2.00 mg/ml, 30 krpm
4.00 mg/ml, 30 krpm
Species class 1
Species class 2
Species class 3
c(s, *) (S-1)
s20°C, w (S)
C
c(s) (S-1)
Interference optics @ 30000 rpm - Signal from all components
D
E
F
Absorbance optics @ 42000 rpm - Signal mainly from RNA
G
Lost material in first run!
Concentration lower than expected.
30000 rpm
42000 rpm
H
0.44 mg/ml, 42 krpm
0.75 mg/ml, 42 krpm
1.09 mg/ml, 42 krpm
I
Interference optics @ 42000 rpm - Signal from all components
J
K
L

Fig. A6

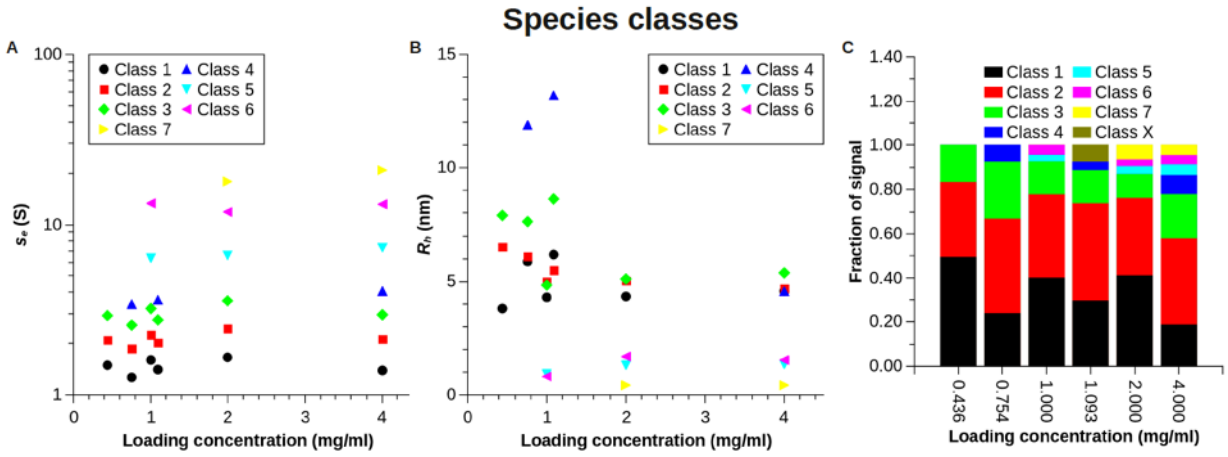


Fig. A7

Table. A3

Results

Species class 1	
Sedimentation coefficient $s^0_{20°C,\,w}$	2.02 [1.61 - 2.43] S
Diffusion coefficient $D^0_{20°C,\,w}$	4.52 [2.76 - 6.28] $\cdot 10^{-7}$ cm²/s
Hydrodynamic radius R_h^0	4.74 [3.41 - 7.74] nm
Species class 2	
Sedimentation coefficient $s^0_{20°C,\,w}$	2.84 [2.28 - 3.41] S
Diffusion coefficient $D^0_{20°C,\,w}$	3.51 [2.87 - 4.15] $\cdot 10^{-7}$ cm²/s
Hydrodynamic radius R_h^0	6.11 ± [5.17 - 7.46] nm
Species class 3	
Sedimentation coefficient $s^0_{20°C,\,w}$	3.96 [2.97 - 4.94] S
Diffusion coefficient $D^0_{20°C,\,w}$	2.92 [1.37 - 4.47] $\cdot 10^{-7}$ cm²/s
Hydrodynamic radius R_h^0	7.34 [4.79 - 15.61] nm

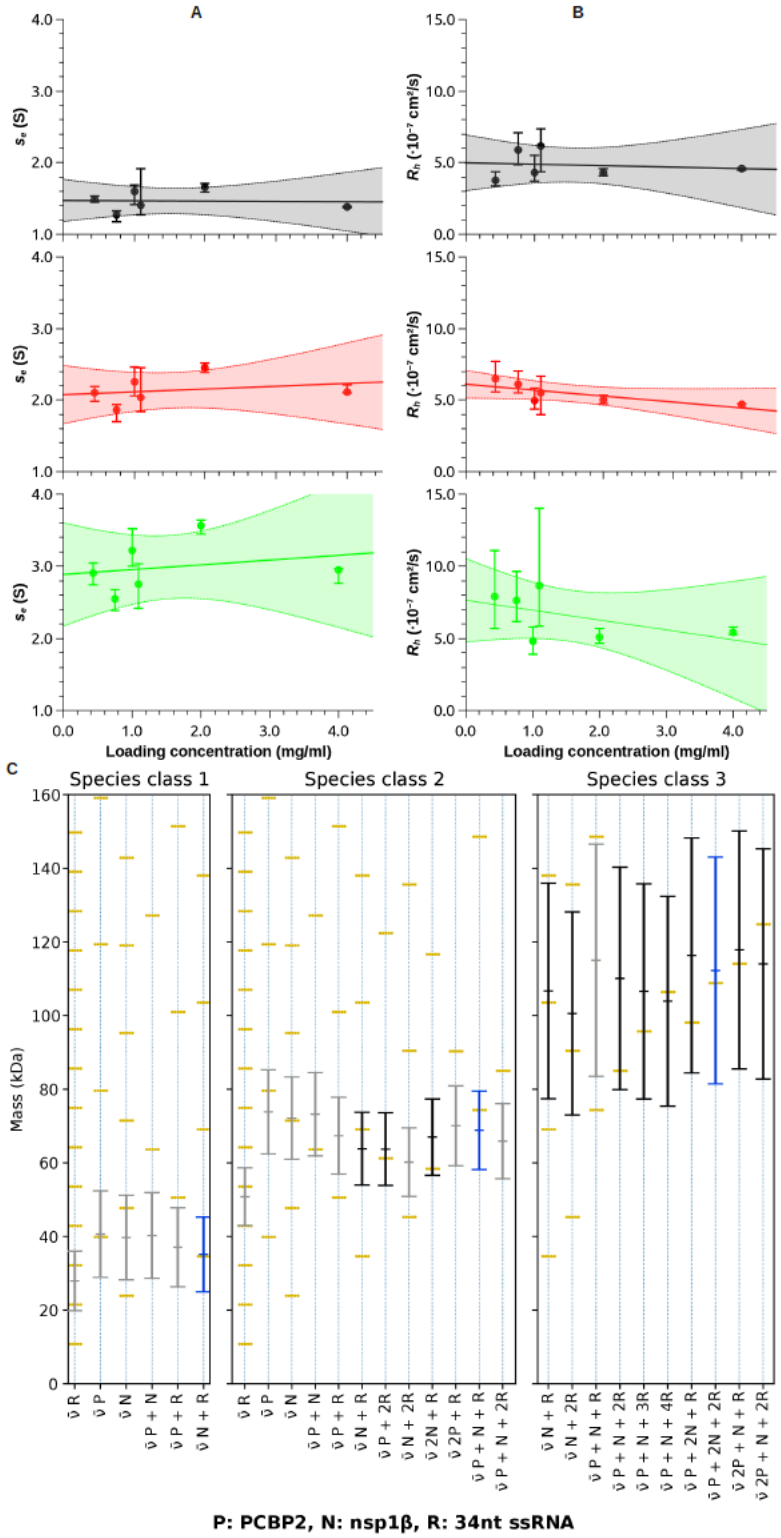

A
B
s_e (S)
R_h ($\cdot 10^{-7}$ cm²/s)
Loading concentration (mg/ml)
C
Species class 1
Species class 2
Species class 3
Mass (kDa)
ν̄ R
ν̄ P
ν̄ N
ν̄ P + N
ν̄ P + R
ν̄ N + R
ν̄ P + 2R
ν̄ N + 2R
ν̄ 2N + R
ν̄ 2P + R
ν̄ P + N + R
ν̄ P + N + 2R
ν̄ N + R
ν̄ N + 2R
ν̄ P + N + R
ν̄ P + N + 2R
ν̄ P + N + 3R
ν̄ P + N + 4R
ν̄ P + 2N + R
ν̄ P + 2N + 2R
ν̄ 2P + N + R
ν̄ 2P + N + 2R
P: PCBP2, N: nsp1β, R: 34nt ssRNA

Fig. A8

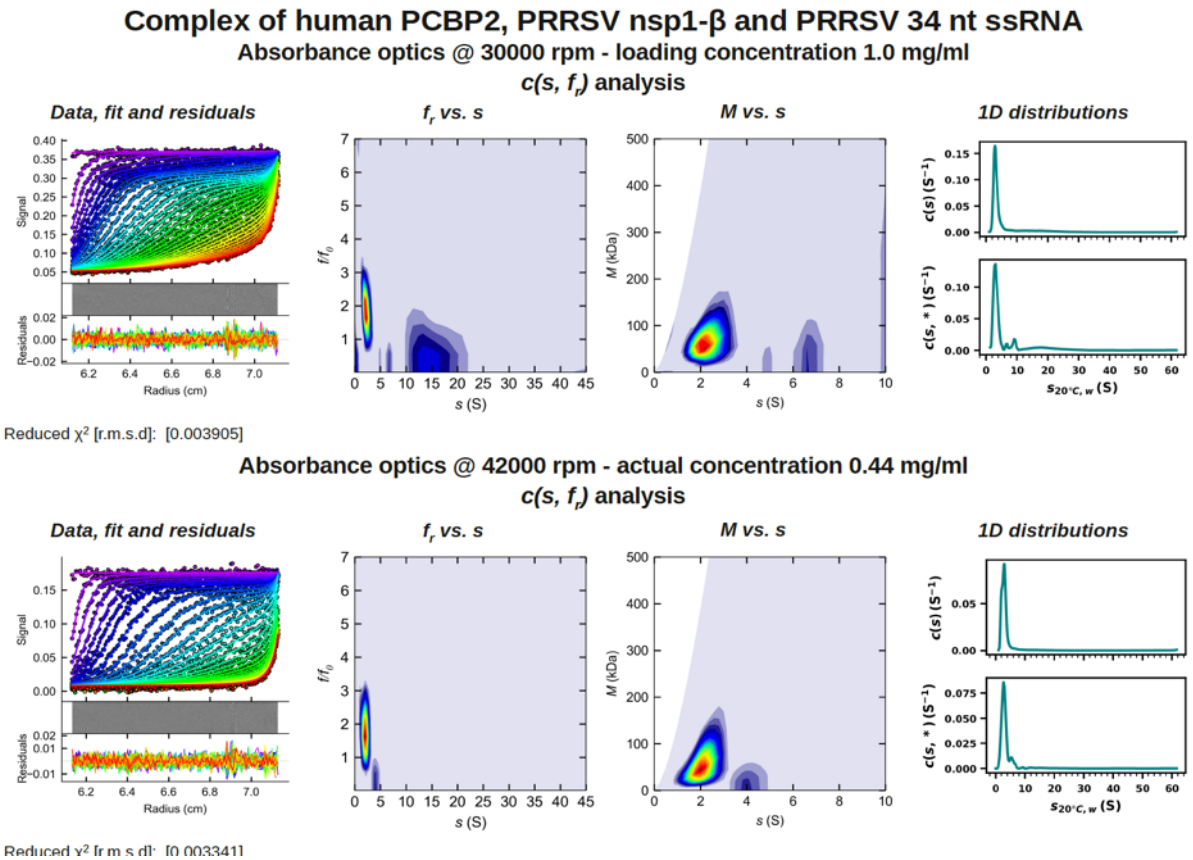

Complex of human PCBP2, PRRSV nsp1-β and PRRSV 34 nt ssRNA
Absorbance optics @ 30000 rpm - loading concentration 1.0 mg/ml
c(s, f_r) analysis
Data, fit and residuals
f_r vs. s
M vs. s
1D distributions
Signal
Residuals
Radius (cm)
f/f_0
s (S)
M (kDa)
c(s) (S^-1)
c(s, *) (S^-1)
s_20°C, w (S)
Reduced χ² [r.m.s.d]: [0.003905]
Absorbance optics @ 42000 rpm - actual concentration 0.44 mg/ml
c(s, f_r) analysis
Data, fit and residuals
f_r vs. s
M vs. s
1D distributions
Reduced χ² [r.m.s.d]: [0.003341]

Fig. A9

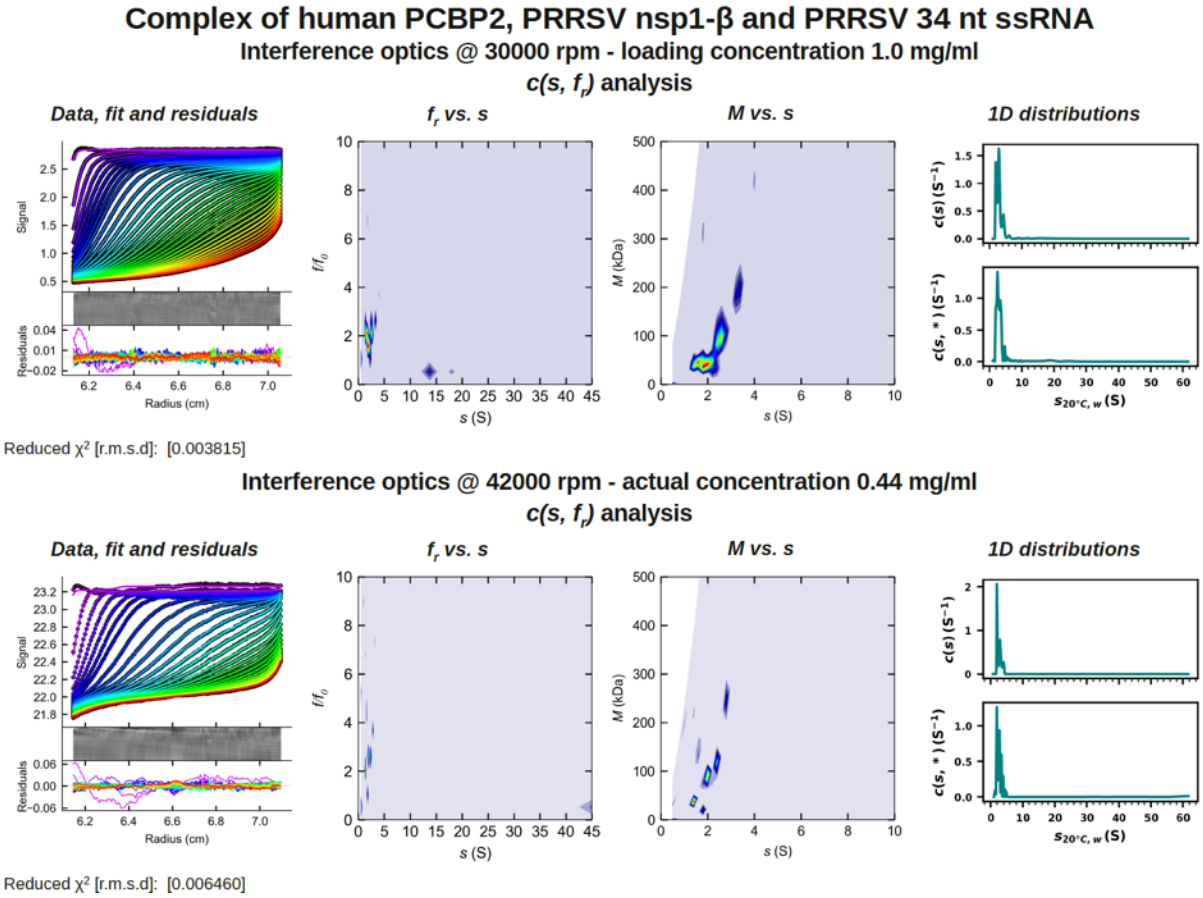

Complex of human PCBP2, PRRSV nsp1-β and PRRSV 34 nt ssRNA
Interference optics @ 30000 rpm - loading concentration 1.0 mg/ml
c(s, f_r) analysis
Data, fit and residuals
f_r vs. s
M vs. s
1D distributions
Signal
Residuals
Radius (cm)
f/f_0
s (S)
M (kDa)
c(s) (S^-1)
c(s, *) (S^-1)
s_20°C, w (S)
Reduced χ² [r.m.s.d]: [0.003815]
Interference optics @ 42000 rpm - actual concentration 0.44 mg/ml
c(s, f_r) analysis
Data, fit and residuals
f_r vs. s
M vs. s
1D distributions
Reduced χ² [r.m.s.d]: [0.006460]

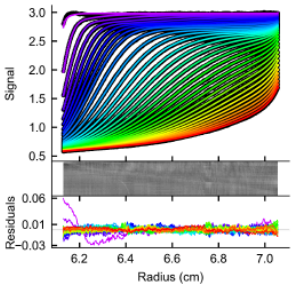
Fig. A10

Complex of human PCBP2, PRRSV nsp1-β and PRRSV 34 nt ssRNA

Species Analysis (1.00 mg/ml, 30000 rpm, i. o.)

Parameter	Value	95% confidence interval
Species # 1 (class 1)		
Sedimentation coefficient $s_{20°C,w}$ (S)	2.185538	1.9320 - 2.3175
Molecular mass[1] *M* (Da)	39429.258626	31975.1983 - 43851.1943
Concentration (signal)	0.889672	
Species # 2 (class 2)		
Sedimentation coefficient $s_{20°C,w}$ (S)	3.086904	2.8131 - 3.3679
Molecular mass[1] *M* (Da)	64508.620238	56946.9871 - 72146.4835
Concentration (signal)	0.848381	
Species # 3 (class 3)		
Sedimentation coefficient $s_{20°C,w}$ (S)	4.413055	4.1166 - 4.8238
Molecular mass[1] *M* (Da)	89133.977232	75536.5637 - 117267.7025
Concentration (signal)	0.327668	
Species # 4 (class 5)		
Sedimentation coefficient $s_{20°C,w}$ (S)	8.680858	6.5740 - 11.0911
Molecular mass[1] *M* (Da)	33004.982556	27014.9568 - 41589.0286
Concentration (signal)	0.0689382	
Species # 5 (class 6)		
Sedimentation coefficient $s_{20°C,w}$ (S)	18.269621	17.2184 - 19.6377
Molecular mass[1] *M* (Da)	62045.164082	50792.5477 - 83330.0656
Concentration (signal)	0.101118	
Common		
Meniscus position (cm)	6.107007	6.1061 - 6.1081
Bottom position (cm)	7.216139	7.2101 - 7.2245
Buffer signal[2] frictional ratio f_r	2.382192	1.8745 - 3.3470
Reduced χ2 of fit [rmsd]	0.1871007	
R.m.s.d. of fit	0.004326	

[1]Substitute $\bar{v} = 0.73$ cm^3/g
[2]Continuous *c(s)* from 0.0 - 0.8 S



Reduced χ^2 [r.m.s.d]: 0.187101 [0.004326]

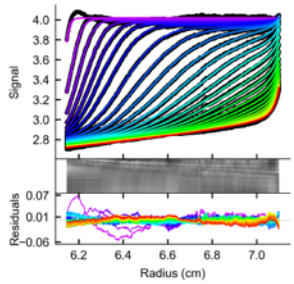
Fig. A11

Complex of human PCBP2, PRRSV nsp1-β and PRRSV 34 nt ssRNA

Species Analysis (0.44 mg/ml, 42000 rpm, i. o.)

Parameter	Value	95% confidence interval
	Species # 1 (class 1)	
Sedimentation coefficient $s_{20°C,w}$ (S)	2.045761	1.9890 - 2.0955
Molecular mass[1] *M* (Da)	32468.185230	28140.1889 - 35903.1468
Concentration (signal)	0.564412	
	Species # 2 (class 2)	
Sedimentation coefficient $s_{20°C,w}$ (S)	2.874757	2.7103 - 2.9979
Molecular mass[1] *M* (Da)	78493.617851	67236.8428 - 93921.0463
Concentration (signal)	0.388871	
	Species # 3 (class 3)	
Sedimentation coefficient $s_{20°C,w}$ (S)	3.976393	3.7610 - 4.1698
Molecular mass[1] *M* (Da)	131871.067920	94586.0501 - 201092.7657
Concentration (signal)	0.193644	
	Common	
Meniscus position (cm)	6.107400	6.1056 - 6.1093
Bottom position (cm)	7.201615	7.1931 - 7.2121
Buffer signal[2] frictional ratio f_r	1.481410	1.2663 - 1.9532
Reduced χ2 of fit [rmsd]	0.5931672	
R.m.s.d. of fit	0.007702	

[1]Substitute $\bar{v}$ = 0.73 cm^3/g
[2]Continuous *c(s)* from 0.0 - 0.8 S



Reduced χ^2 [r.m.s.d]: 0.593167 [0.007702]

Fig. A12

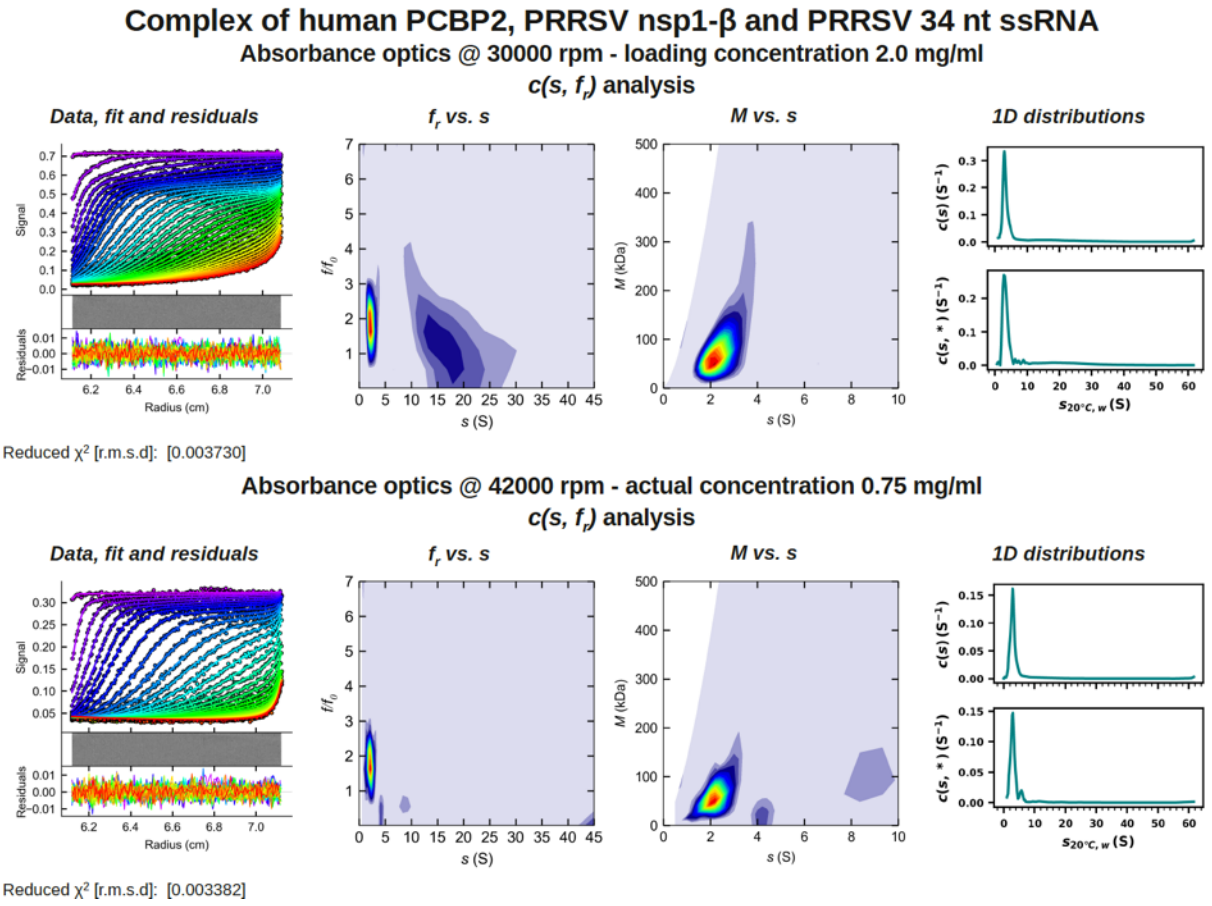

Complex of human PCBP2, PRRSV nsp1-β and PRRSV 34 nt ssRNA
Absorbance optics @ 30000 rpm - loading concentration 2.0 mg/ml
c(s, f_r) analysis
Data, fit and residuals
f_r vs. s
M vs. s
1D distributions
Signal
Residuals
Radius (cm)
f/f_0
s (S)
M (kDa)
c(s) (S^-1)
c(s, *) (S^-1)
s_20°C, w (S)
Reduced χ² [r.m.s.d]: [0.003730]
Absorbance optics @ 42000 rpm - actual concentration 0.75 mg/ml
c(s, f_r) analysis
Data, fit and residuals
f_r vs. s
M vs. s
1D distributions
Reduced χ² [r.m.s.d]: [0.003382]

Fig. A13

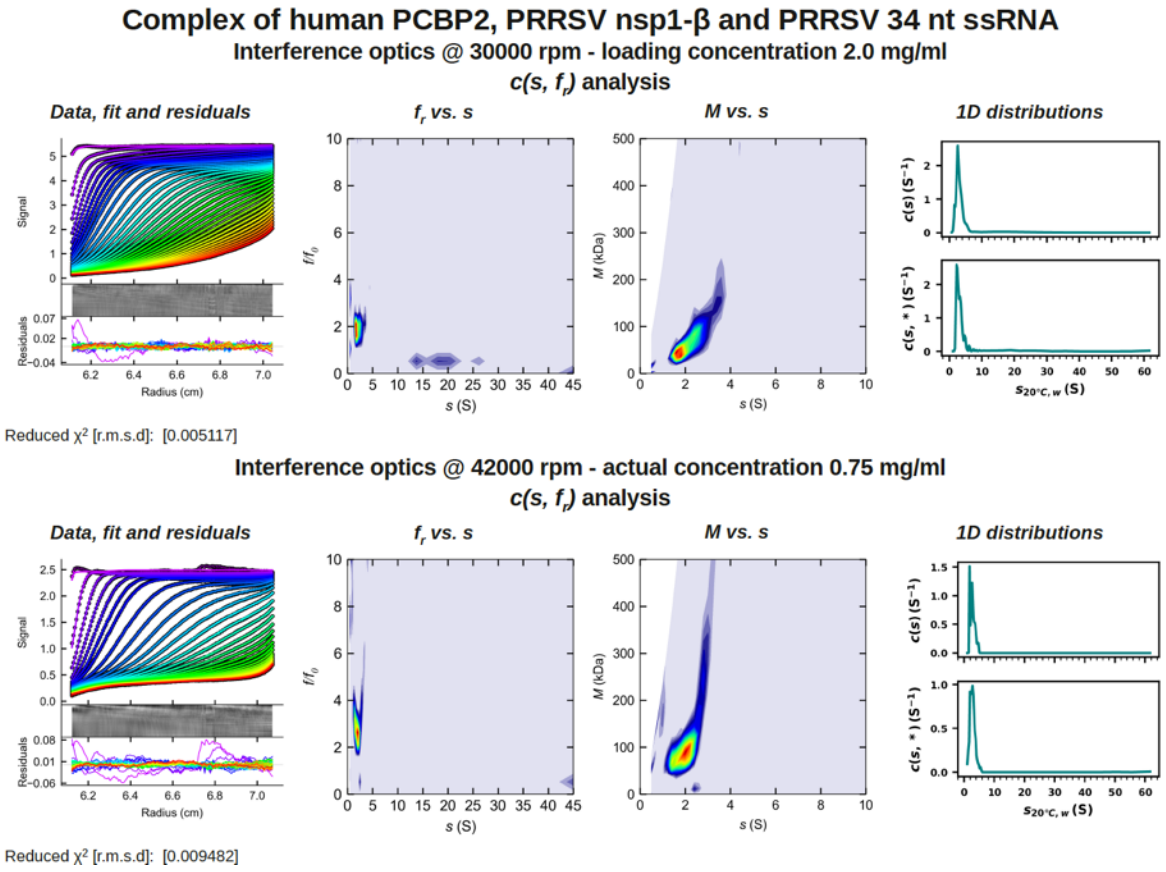

Complex of human PCBP2, PRRSV nsp1-β and PRRSV 34 nt ssRNA
Interference optics @ 30000 rpm - loading concentration 2.0 mg/ml
c(s, f_r) analysis
Data, fit and residuals
f_r vs. s
M vs. s
1D distributions
Signal
Residuals
Radius (cm)
f/f_0
s (S)
M (kDa)
c(s) (S^-1)
c(s, *) (S^-1)
s_20°C, w (S)
Reduced χ² [r.m.s.d]: [0.005117]
Interference optics @ 42000 rpm - actual concentration 0.75 mg/ml
c(s, f_r) analysis
Data, fit and residuals
f_r vs. s
M vs. s
1D distributions
Reduced χ² [r.m.s.d]: [0.009482]

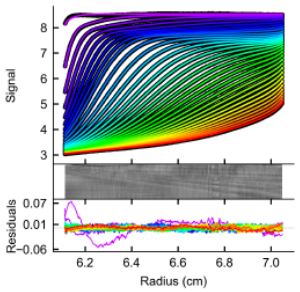
Fig. A14

Complex of human PCBP2, PRRSV nsp1-β and PRRSV 34 nt ssRNA

Species Analysis (2.00 mg/ml, 30000 rpm, i. o.)

Parameter	Value	95% confidence interval
Species # 1 (class 1)		
Sedimentation coefficient $s_{20°C,w}$ (S)	2.259298	2.1789 - 2.3345
Molecular mass[1] *M* (Da)	40911.383621	39015.7993 - 42569.8757
Concentration (signal)	2.10665	
Species # 2 (class 2)		
Sedimentation coefficient $s_{20°C,w}$ (S)	3.362426	3.2647 - 3.4499
Molecular mass[1] *M* (Da)	70607.408753	66954.0626 - 74408.5887
Concentration (signal)	1.80079	
Species # 3 (class 3)		
Sedimentation coefficient $s_{20°C,w}$ (S)	4.879189	4.7237 - 4.9818
Molecular mass[1] *M* (Da)	104525.361816	94748.6103 - 115580.2252
Concentration (signal)	0.557564	
Species # 4 (class 5)		
Sedimentation coefficient $s_{20°C,w}$ (S)	9.007732	7.8572 - 10.1107
Molecular mass[1] *M* (Da)	49129.251609	38405.4058 - 69148.1903
Concentration (signal)	0.179173	
Species # 5 (class 6)		
Sedimentation coefficient $s_{20°C,w}$ (S)	16.172510	15.2395 - 17.1308
Molecular mass[1] *M* (Da)	114826.491328	86165.0080 - 173746.5190
Concentration (signal)	0.15973	
Species # 6 (class 7)		
Sedimentation coefficient $s_{20°C,w}$ (S)	24.574698	23.3895 - 25.9063
Molecular mass[1] *M* (Da)	42732.956296	38685.9354 - 49078.2030
Concentration (signal)	0.337733	
Common		
Meniscus position (cm)	6.078588	6.0778 - 6.0790
Bottom position (cm)	7.233989	7.2288 - 7.2399
Buffer signal[2] frictional ratio *f*$_r$	3.336644	2.8806 - 3.9916
Reduced χ2 of fit [rmsd]	0.3392895	
R.m.s.d. of fit	0.005825	

[1]Substitute $\bar{v}$ = 0.73 cm^3/g
[2]Continuous *c(s)* from 0.0 - 0.8 S



Reduced χ^2 [r.m.s.d]: 0.339290 [0.005825]

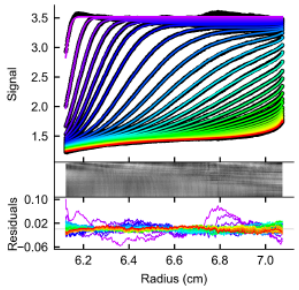
Fig. A15

Complex of human PCBP2, PRRSV nsp1-β and PRRSV 34 nt ssRNA

Species Analysis (0.75 mg/ml, 42000 rpm, i. o.)

Parameter	Value	95% confidence interval
Species # 1 (class 1)		
Sedimentation coefficient $s_{20°C,w}$ (S)	1.737800	1.6152 - 1.8163
Molecular mass[1] *M* (Da)	42836.913056	36022.7457 - 53638.9211
Concentration (signal)	0.488908	
Species # 2 (class 2)		
Sedimentation coefficient $s_{20°C,w}$ (S)	2.545311	2.3295 - 2.6511
Molecular mass[1] *M* (Da)	65269.106310	58823.6621 - 72195.2814
Concentration (signal)	0.869682	
Species # 3 (class 3)		
Sedimentation coefficient $s_{20°C,w}$ (S)	3.497679	3.2748 - 3.6633
Molecular mass[1] *M* (Da)	112116.879536	90117.1719 - 141988.8079
Concentration (signal)	0.526914	
Species # 4 (class 4)		
Sedimentation coefficient $s_{20°C,w}$ (S)	4.655214	4.3692 - 4.8847
Molecular mass[1] *M* (Da)	232327.712281	148281.3925 - 426686.8809
Concentration (signal)	0.156663	
Common		
Meniscus position (cm)	6.093045	6.0911 - 6.0950
Bottom position (cm)	7.175154	7.1615 - 7.1904
Buffer signal[2] frictional ratio *f*,	5.718135	4.3407 - 7.4711
Reduced χ2 of fit [rmsd]	0.9201238	
R.m.s.d. of fit	0.009592	

[1]Substitute $\bar{\nu}$ = 0.73 cm^3/g
[2]Continuous *c(s)* from 0.0 - 0.8 S



Reduced χ^2 [r.m.s.d]: 0.920124 [0.009592]

Fig. A16

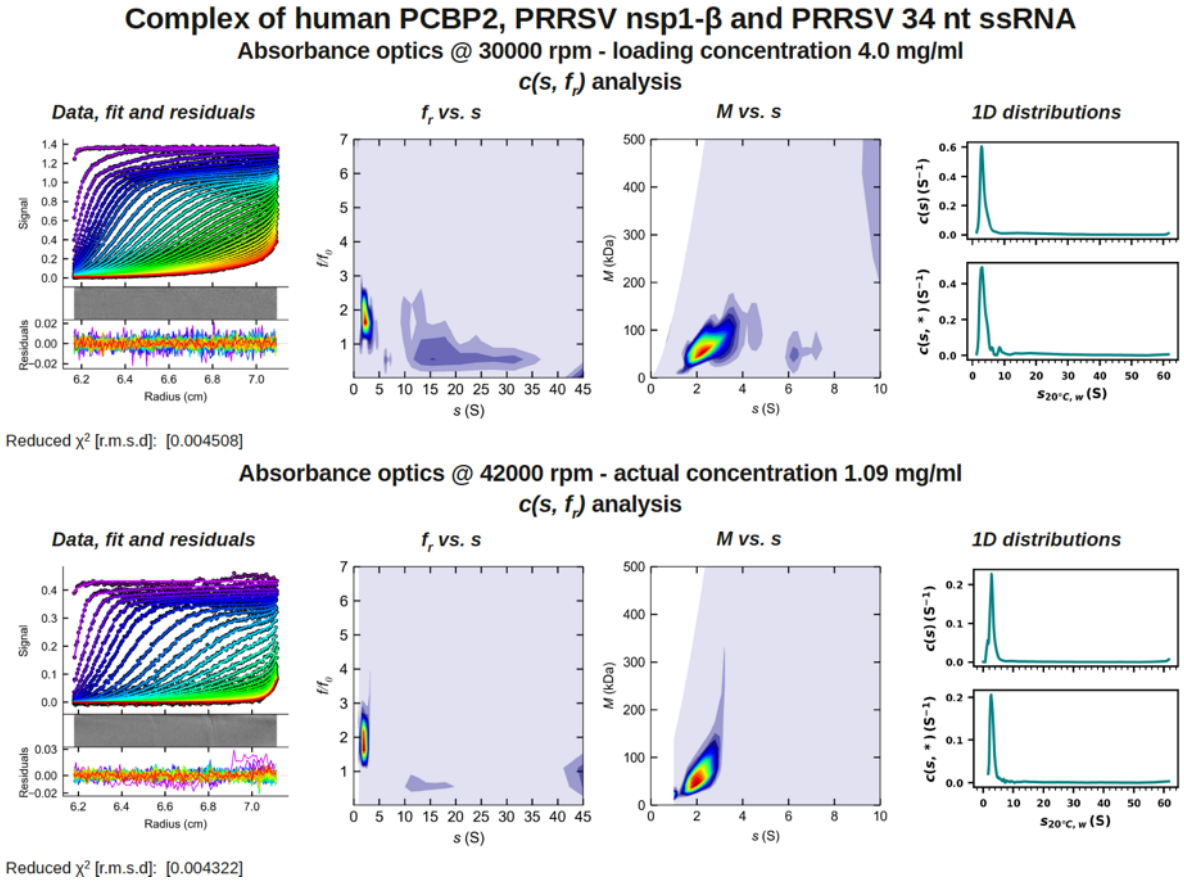

Complex of human PCBP2, PRRSV nsp1-β and PRRSV 34 nt ssRNA
Absorbance optics @ 30000 rpm - loading concentration 4.0 mg/ml
c(s, f_r) analysis
Data, fit and residuals
f_r vs. s
M vs. s
1D distributions
Signal
Residuals
Radius (cm)
f/f_0
s (S)
M (kDa)
c(s) (S^-1)
c(s, *) (S^-1)
s_20°C, w (S)
Reduced χ² [r.m.s.d]: [0.004508]
Absorbance optics @ 42000 rpm - actual concentration 1.09 mg/ml
c(s, f_r) analysis
Data, fit and residuals
f_r vs. s
M vs. s
1D distributions
Reduced χ² [r.m.s.d]: [0.004322]

Fig. A17

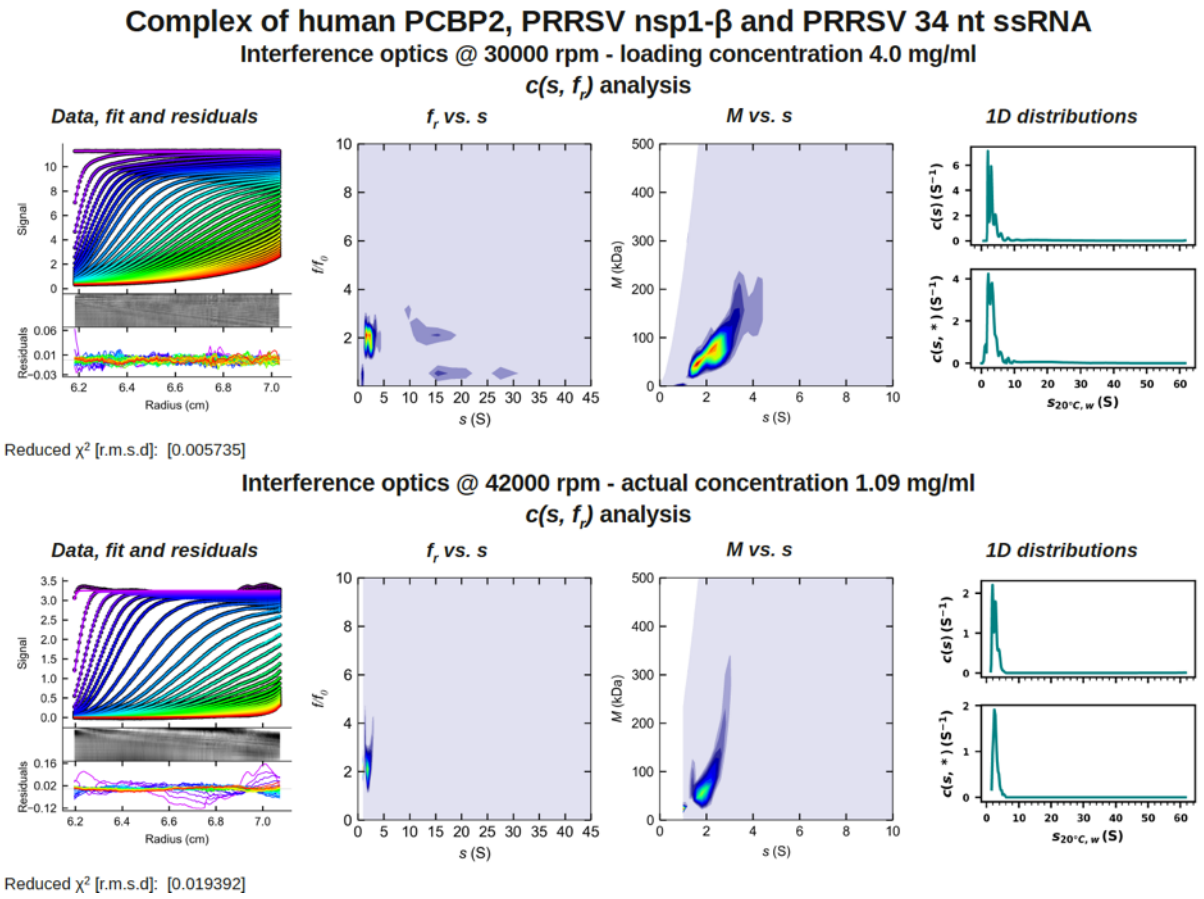

Complex of human PCBP2, PRRSV nsp1-β and PRRSV 34 nt ssRNA
Interference optics @ 30000 rpm - loading concentration 4.0 mg/ml
c(s, f_r) analysis
Data, fit and residuals
f_r vs. s
M vs. s
1D distributions
Signal
Residuals
Radius (cm)
f/f_0
s (S)
M (kDa)
c(s) (S^-1)
c(s, *) (S^-1)
s_20°C, w (S)
Reduced χ² [r.m.s.d]: [0.005735]
Interference optics @ 42000 rpm - actual concentration 1.09 mg/ml
c(s, f_r) analysis
Data, fit and residuals
f_r vs. s
M vs. s
1D distributions
Reduced χ² [r.m.s.d]: [0.019392]

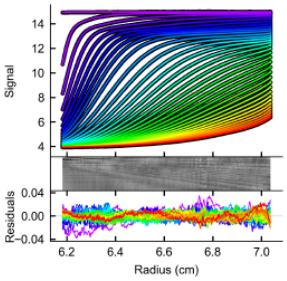
Fig. A18

Complex of human PCBP2, PRRSV nsp1-β and PRRSV 34 nt ssRNA

Species Analysis (4.00 mg/ml, 30000 rpm, i. o.)

Parameter	Value	95% confidence interval
Species # 1 (class 1)		
Sedimentation coefficient $s_{20°C,w}$ (S)	1.893838	1.8767 - 1.9115
Molecular mass[1] *M* (Da)	36242.513823	35605.9184 - 36924.8539
Concentration (signal)	2.02274	
Species # 2 (class 2)		
Sedimentation coefficient $s_{20°C,w}$ (S)	2.895957	2.8717 - 3.0216
Molecular mass[1] *M* (Da)	56798.604800	56171.7866 - 57417.8843
Concentration (signal)	4.20148	
Species # 3 (class 3)		
Sedimentation coefficient $s_{20°C,w}$ (S)	4.032742	3.7914 - 4.0755
Molecular mass[1] *M* (Da)	91110.192293	88642.3598 - 93600.9839
Concentration (signal)	2.17643	
Species # 4 (class 4)		
Sedimentation coefficient $s_{20°C,w}$ (S)	5.558437	5.5059 - 5.7911
Molecular mass[1] *M* (Da)	106503.214501	97771.9841 - 113202.5506
Concentration (signal)	0.924506	
Species # 5 (class 5)		
Sedimentation coefficient $s_{20°C,w}$ (S)	9.980081	9.6854 - 10.2378
Molecular mass[1] *M* (Da)	55993.213184	50533.5330 - 65497.6998
Concentration (signal)	0.504232	
Species # 6 (class 6)		
Sedimentation coefficient $s_{20°C,w}$ (S)	18.163020	17.9089 - 18.4664
Molecular mass[1] *M* (Da)	116367.123008	102061.2320 - 138132.1828
Concentration (signal)	0.460625	
Species # 7 (class 7)		
Sedimentation coefficient $s_{20°C,w}$ (S)	28.605184	27.7262 - 30.0465
Molecular mass[1] *M* (Da)	50122.522087	49947.7361 - 50312.8130
Concentration (signal)	0.500927	
Common		
Meniscus position (cm)	6.154459	6.1540 - 6.1548
Bottom position (cm)	7.219491	7.2170 - 7.2245
Buffer signal[2] frictional ratio f_r	1.126869	0.9060 - 1.3521
Reduced χ2 of fit [rmsd]	0.5273897	
R.m.s.d. of fit	0.007262	

[1]Substitute $\bar{\nu}$ = 0.73 cm^3/g
[2]Continuous *c(s)* from 0.0 - 0.8 S



Reduced χ^2 [r.m.s.d]: 0.527390 [0.007262]

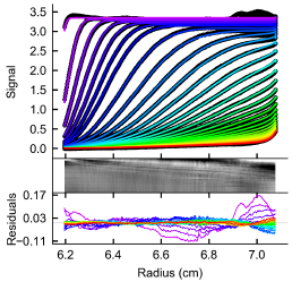
Fig. A19

Complex of human PCBP2, PRRSV nsp1-β and PRRSV 34 nt ssRNA

Species Analysis (1.09 mg/ml, 42000 rpm, i. o.)

Parameter	Value	95% confidence interval
Species # 1 (class X)		
Sedimentation coefficient $s_{20°C,w}$ (S)	1.209722	0.8561 - 1.7601
Molecular mass[1] *M* (Da)	30400.731201	20063.1666 - >900000
Concentration (signal)	0.255537	
Species # 2 (class 1)		
Sedimentation coefficient $s_{20°C,w}$ (S)	1.922069	1.7474 - 2.6281
Molecular mass[1] *M* (Da)	49646.909726	42877.4221 - 61527.1873
Concentration (signal)	0.962814	
Species # 3 (class 2)		
Sedimentation coefficient $s_{20°C,w}$ (S)	2.787625	2.5211 - 3.3593
Molecular mass[1] *M* (Da)	64072.302750	54843.6506 - 109988.7832
Concentration (signal)	1.43393	
Species # 4 (class 3)		
Sedimentation coefficient $s_{20°C,w}$ (S)	3.776941	3.3105 - 4.1561
Molecular mass[1] *M* (Da)	136947.108751	87118.2414 - 219814.9212
Concentration (signal)	0.4912	
Species # 5 (class 4)		
Sedimentation coefficient $s_{20°C,w}$ (S)	4.936165	4.2947 - 5.5977
Molecular mass[1] *M* (Da)	273002.737766	108529.5424 - >900000
Concentration (signal)	0.119144	
Common		
Meniscus position (cm)	6.169227	6.1674 - 6.1710
Bottom position (cm)	7.181567	7.1682 - 7.2017
Buffer signal[2] frictional ratio f_r	4.608594	2.9507 - 28.0506
Reduced χ2 of fit [rmsd]	3.344057	
R.m.s.d. of fit	0.018287	

[1]Substitute $\bar{v}$ = 0.73 cm^3/g
[2]Continuous *c(s)* from 0.0 - 0.8 S



Reduced χ^2 [r.m.s.d]: 3.344057 [0.018287]

Fig. A20

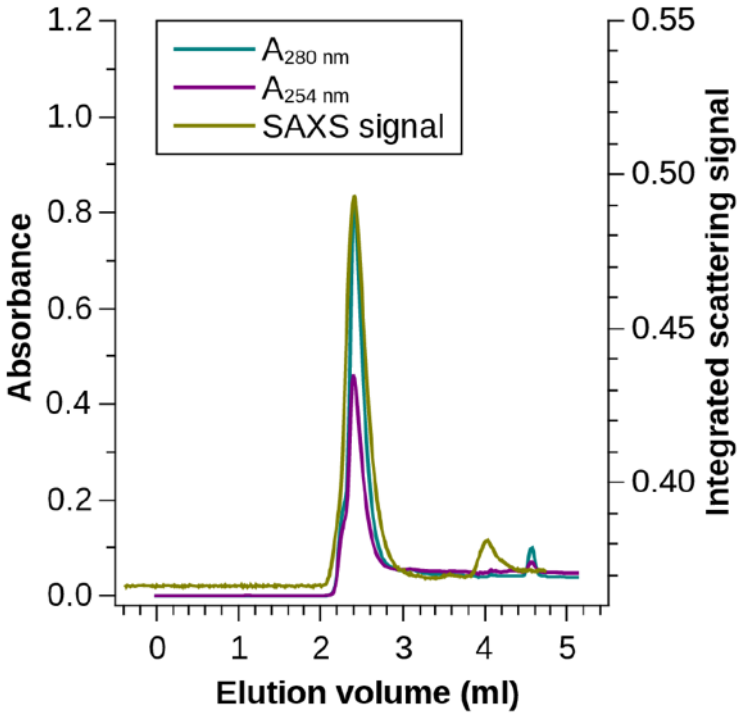

A280 nm
A254 nm
SAXS signal
Absorbance
Integrated scattering signal
Elution volume (ml)
1.2
1.0
0.8
0.6
0.4
0.2
0.0
0.55
0.50
0.45
0.40
0
1
2
3
4
5

Fig. A21

3 Chapter 3: Investigating the protease and deubiquitinating activity of members of genus Marafivirus with aims to develop resilient transgenics

A majority of this research was originally published in the *Journal of Biological Chemistry*.

Patel, A., McBride, J.A.M., Mark, B.L.. The endopeptidase of the maize-affecting Marafivirus type member maize rayado fino virus doubles as a deubiquitinase. *J Biol Chem*. 2021. doi: 10.1016/j.jbc.2021.100957.

Contributions statement:

The overall conceptualization of this project was done by B.L.M. Many of the experiments were conceptualized by A.P. Ubiquitin variant phage display library work (selections) was done by members of Dr. Sachdev Sidhu's lab (Drs. Wei Zhang and Jacky Chung, U of Toronto). Maize transgenic development (excluding tDNA plasmid construction) was done by the Plant Transformation Facility at the Iowa State University and members at the USDA in Fargo, ND (Drs. John Weiland and Melvin Bolton). All remaining work was done by A.P. with the assistance of undergraduate researchers Oxana Katova, Jessica McBride and Nathan Glowach.

3.1 *Introduction*

As above mentioned, positive-sense single-stranded RNA (+ssRNA) viruses have evolved remarkable polycistronic translational mechanisms that maximize genomic coding capacity to produce the viral proteins needed for replication and packaging (10, 235). The expression of polyproteins allows viruses to produce a full set of proteins from their genomes without needing to encode for additional genetic features that would typically direct and regulate translation. +ssRNA viruses infect virtually all forms of life (57). Significant attention has been paid to human and animal-affecting +ssRNA viruses that are capable of cross-species transmission; however, the +ssRNA virome affecting plants appears to be far more diverse and abundant (57, 236). With the threat of food shortage becoming more of a reality in a matter of decades due to climate change and the quickly increasing global population (63, 147), understanding the relationship between crops important to global food security and their viral pathogens will aid in developing sustainable agricultural practises.

Marafiviruses are +ssRNA viruses, which cause persistent infection in diverse plant species that are agriculturally relevant to food and biofuels production (146, 153, 237). Currently there are ten classified members of the genus *Marafivirus* (141) and additional marafivirus candidates that affect different plant species have been recently identified (135, 136). Marafiviruses are persistent propagative viruses meaning that they are also able to replicate in the leafhopper insect vectors that transmit them to their final plant host (137). Marafiviruses have a single-stranded, monopartite genome with an average size of 6-7 kb, which is translated into a single polyprotein that is roughly 2,000 amino acids in length (Fig. 3.1). The polyprotein characteristically contains four non-structural domains that comprise the viral replicase (methyltransferase (MT), protease (PRO), helicase (HEL) and RNA-dependent RNA polymerase (RdRp)) as well as the major structural coat

protein (CP) for genome packing (238). *Maize rayado fino virus* (MRFV) is the type member of the genus *Marafivirus* within the family *Tymoviridae* (146). The virus was first described in 1969 from Costa Rica and has since been found as far south as Brazil and as far north as the United States (146, 239, 240). MRFV infection causes reduction in plant height and ear development with few to no seeds within certain corn species and can lead to severe agricultural loss with decimation to nearly 100 % in some cultivars (145).

MRFV shares remarkable sequence similarity to the tymovirus type member, *Turnip yellow mosaic virus*; TYMV (~43 % at the amino acid level); however, the amount of foundational research that exists between the two is disparate, with TYMV being far better studied. The papain-like cysteine protease of TYMV (TYMV PRO) is currently the most extensively characterized plant deubiquitinase (149, 163, 172, 173, 241). Over 67 % of plant viral proteases, like all the marafivirus proteases, are cysteine proteases (23). PRO's primary function is to act as a polyprotein processor, but it has been shown to have auxiliary function as a ubiquitin (Ub) hydrolase to aid in bypassing the host innate immune system by removing ubiquitin from the RdRp, preventing its degradation by the ubiquitin proteasome (149, 167, 172, 173). Notably, the junctions between the replicase domains within the marafivirus polyprotein (Fig. 3.1, arrows) closely resemble the C-terminal ^{73}LRGG^{76} tail of ubiquitin (the cleavage site for ubiquitin hydrolases) (168), suggesting marafivirus endopeptidases may also have deubiquitinating activity.

Here we provide the first structural and functional insight into the endopeptidases from the genus *Marafivirus.* We present data illustrating that the PRO domains from six different species within this genus all have *bona fide* deubiquitinating (DUB) activity. We compare the differences and similarities of all six proteases towards recognition of different types of ubiquitin substrates

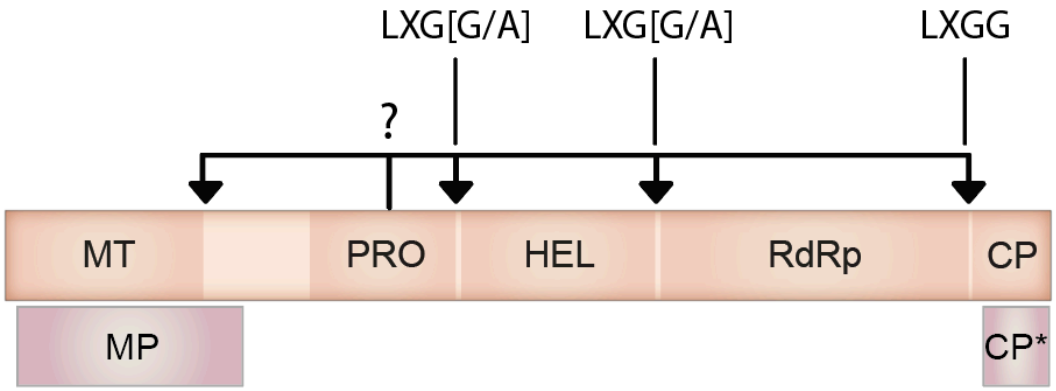


Figure 3.1 Marafivirus polyprotein arrangement and endopeptidase comparison

Simplified schematic of the typical arrangement of the sole polyprotein produced by marafiviruses and two additional proteins produced from genomic RNA. (MT) methyl transferase, (PRO) endopeptidase/deubiquitinase, (HEL) helicase, (RdRp) RNA-dependent RNA polymerase, (CP) major coat protein, (MP) movement protein, (CP*) minor coat protein. Arrows indicate junction points between each domain and the possible cleavage sites targeted by the endopeptidase. The four amino acid residues at each putative cleavage sites are indicated where "X" is any amino acid.

and contrast our findings with those of the tymoviruses. We also discuss the endopeptidase activity of MRFV PRO that is required for polyprotein processing to generate the viral replicase complex. X-ray structures of MRFV PRO alone and bound to Ub reveal unique structural characteristics that differ from TYMV PRO and notable conformational changes that occur in response to the Ub substrate. Together, our data provide new insight into how marafiviruses generate their replicase machinery through polyprotein processing and that they exhibit deubiquitinase activity that may corrupt the immune response of their hosts.

Further, the overarching goal of this chapter remains to exploit the weak Ub-binding tendencies of these viral proteases by utilizing highly selective, synthetic ubiquitin variants (UbV), which ideally would act as competitive inhibitors towards these proteases to stifle polyprotein processing and significantly reduce the effects of viral infection in transgenic plants expressing the UbV transgene.

3.2 Materials and Methods

3.2.1 Primers used in DNA constructs

Table 3.1 Oligos used for all DNA constructs

Oligo name	Sequence (5' to 3')	Purpose
CSDaV_pGEX_F	GATATA**GGATCC**TCCGATTGGGACCC CTCTC	CSDaV PRO into pGEX-6p-1
CSDaV_pGEX_R	TATATC**CTCGAG**TTAACGAGGGGTGG CGCTC	"
MRFV_pGEX_F	GATATA**GGATCC**CCGGAGCCTGACAC GGCC	MRFV PRO into pGEX-6p-1
MRFV_pGEX_R	TATATC**CTCGAG**TTAAAGTGAGAAAT TATCGGC	"
MRFV PRO_{C61A}-$HEL_{N\text{-}term\ Domain}$_F	GATA TA**CATATG**CCGGAACCCGATACC	MRFV PRO_{C61A}-$HEL_{N\text{-}term\ Domain}$ into pET19b
MRFV PRO_{C61A}-$HEL_{N\text{-}term\ Domain}$_R	TATATC**GGATCC**TTAGCAATAA AAGTCTACATAGG	"
MRFV PRO-$HEL_{N\text{-}term\ Domain}$_F	(P)-CCG TGCCGCTTGCTTACTGGTCG	Mutagen. of MRFV PRO_{C61A}-$HEL_{N\text{-}term\ Domain}$ into pET19b to WT
MRFV PRO-$HEL_{N\text{-}term\ Domain}$_R	(P)-GTTGGATAAGGGATAGAG	"
TYMV_pGEX_F	GATATA**GGATCC**GGCTCCTCACAATT ACTGCC	TYMV PRO into pGEX-6p-1
TYMV_pGEX_R	TATATC**CTCGAG**TTAAGAGCCTAAGA GACGTTTACCC	"
CSDaV_pHBT95_F	GATATA**GGATCC**ATGTCCGATTGGGA CC	CSDaV PRO into pHBT95
CSDaV_pHBT95_R	TATATC**GCGGCCGC**TTTAA GTTGAATC	"
$CSDaV_{C56A}$_pHBT95_F	(P)-ACTGTTGTCTATTGAACAAGCG	Mutagen. of CSDaV PRO_{C56A} into pHBT95
$CSDaV_{C56A}$_pHBT95_R	(P)-AGAGCATCCATGTCTGGGTAG	"
CSDaV_PROHEL_pHBT95_F	GATATA**GGATCC**ATGTATCCATACG	CSDaV PROHEL into pHBT95
CSDaV_PRO_{C56A}HEL_pHBT95_R	TATATC**GCGGCCGC**TTTACTTG	"

CSDaV_PRO$_{C56Apro}$HEL_pHBT95_F	P-ACTGTTGTCTATTGAACAAGCG	Mutagen. of CSDaV_PRO$_{C56A}$HEL in pHBT95
CSDaV_PROHEL_pHBT95_R	P-AGAGCATCCATGTCTGGGTAG	"
UbAA_pΩ_F	(P)-AGGGGATCCTCTAGAGTCGAC	Mutagen of Ub ^{75}GG^{76} to AA in pΩ
UbAA_pΩ_R	(P)-CATGCTGCTCTTAGTCTTAAGACA	"
CSDaV_ubv_ pΩ_F	GATATA**CCATGG**AACAAAAATTG	CSDaV UbV into pΩ
CSDaV_ubv_ pΩ_R	TATATC**GGATCC**CCTCAAGTTCC	"
MRFVUbv2_pDW3596F	GATATA**GGATCC**GATCCAAATCCG	MRFV UbV2 into pDW3596
MRFVUbv2_pDW3596R	TATATC**ACTAGT**TCATCAATGGACTGG	"

*Restriction sites in bold
*(P) indicates 5' phosphorylation

3.2.2 DNA constructs

3.2.2.1 DNA constructs for recombinant protein expression and purification

Synthetic DNA (Integrated DNA technologies) coding for open reading frames of PRO domains from MRFV, TYMV and CSDaV (*Citrus sudden death associated virus*) were amplified by PCR using primers listed in Table 3.1 and ligated into pGEX-6P-1 (GE Healthcare) using BamHI and XhoI restriction sites. The remaining PRO open reading frames from BlVS (*Blackberry virus S*), GSyV1 (*Grapevine Syrah virus 1*), OLV3 (*Olive latent virus 3*) and OBDV (*Oat blue dwarf virus*) were constructed by GenScript using the same cloning strategy. NCBI reference sequences for each virus: (BlVS NC_038328; CSDaV NC_006950; GSyV1 NC_002786, OLV3 NC_013920; OBDV NC_001793; TYMV NC_004063).

The MRFV PRO$_{C61A}$-HEL$_{N\text{-term Domain}}$ fusion protein expression construct (residues 667-1038 from its polyprotein; UniProtKB-Q91TW9) was generated from codon-optimized synthetic

DNA (Integrated DNA technologies) and PCR using primers listen in Table 3.1. The amplicon was cloned into pET19b (Novagen) in frame with the native N-terminal His_{10} tag using NdeI and BamHI restriction sites. The WT version of this construct (MRFV PRO-$HEL_{N\text{-term Domain}}$), which retains the active site cysteine of PRO was designed using phosphorylated primers (Table 3.1) and site-directed mutagenesis to reintroduce the cysteine that was formerly an alanine using the pET19b vector containing C61A mutation as template.

Vectors encoding UbV genes in pET53-DEST were provided by members of Dr. Sachdev Sidhu's lab (U of Toronto).

3.2.2.2 DNA constructs for Arabidopsis transient gene expression assays and maize transformation

The CSDaV PRO expression plasmid (pHBT95-CSdaV_PRO) used for transient V5-tagged PRO expression in *A. thaliana* protoplasts was designed using the pHBT95-sGFP(S65T) vector (kindly provided by Dr. Jen Sheen, Dept. of Genetics, Harvard Medical School). The GFP coding sequence was excised using BamHI and NotI restriction enzymes and replaced with the coding sequence representing CSDaV PRO, which was synthetically derived, and codon optimized for expression in *A. thaliana* (Integrated DNA technologies) using primers listed in Table 3.1. HBT95-CSDaV_PRO_{C56A} was generated using phosphorylated primers (Table 3.1) and 'round-the-horn PCR (242) using HBT95-CSDaV_PRO as template. Constructs encoding CSDaV HA-PROHEL-FLAG were also prepared using synthetic DNA and primers listed in Table 3.1 and was ligated into pHBT95 identically to above. The catalytic mutant (HA-PRO_{C56A}HEL-FLAG mutant) was constructed using primers in Table 3.1 and site-directed mutagenesis PCR, using the WT plasmid as template.

The construct that was used to transiently express double Myc-tagged Ubiquitin in *A. thaliana* (pΩ-$(Myc)_2$) was kindly provided by Dr. Isabelle Jupin (Dept. of Molecular Virology, Institut Jacques Monod). Site-directed mutagenesis using phosphorylated primers (Table 3.1) were used to generate a mutant plasmid in which the terminal glycine residues (^{75}GG^{76}) were altered to alanine. The construct encoding for CSDaV UbV1 was prepared from synthetic DNA, PCR amplification (Table 3.1 for primers) and ligated into pΩ-$(Myc)_2$ using BamHI and NotI restriction sites.

A transfer DNA plasmid was generated to include the MRFV UbV2 transgene for *Agrobacterium*-mediated transformation of *Zea mays* (line B104). The empty plasmid (pDW3596) was sent to us by the Iowa State University Plant Transformation Facility. As per the facility's recommendation, the Rice U6 promoter and TonB genes were removed from the plasmid with SacI and EcoRI restriction sites and was then re-circularized with two complementary, 5' phosphorylated oligos that had mimicking SacI/EcoRI overhangs. Synthetic DNA coding for an N-term 3xFLAG tag and the MRFV UbV2 was codon optimized for *Z. mays* and inserted into pDW3596 using BamHI and SpeI restriction sites in between the right and left borders of the binary vector.

3.2.3 Protein expression and purification

Expression plasmids for the marafivirus PRO domains (and TYMV PRO) were used to transform *E. coli* BL21 (DE3) GOLD cells (Stratagene) for protein production. Transformed *E. coli* were grown overnight at 37 °C in lysogeny broth (LB) containing 150 μg/ml ampicillin. The overnight culture was then used to inoculate 500 ml or 1 L of fresh ampicillin-containing LB (1:50 dilution) and was subsequently grown at 37 °C with shaking to an A_{600} of 0.7-0.8. Expression of

the GST-tagged PRO enzymes (from pGEX-6P-1 constructs) or His_{10}-tagged MRFV PRO_{C61A}-$HEL_{N\text{-}termDomain}$ was induced by the addition of 0.5 mM isopropyl β-D-1 thiogalatopyranoside and left to incubate with shaking at 16 °C for an additional 18 h. Cells were then pelleted by centrifugation and either immediately used or stored at -80 °C.

All PRO domains were purified as follows. Cell pellets were resuspended in ice-cold lysis buffer (50 mM TRIS-HCl pH 8.0, 300 mM NaCl and 2 mM DTT) and lysed using an Avestin Emulsiflex C3 high-pressure cell homogenizer (ATA Scientific Instruments). Cell lysates were clarified by centrifugation (17,211 X g at 4 °C), and supernatant containing GST-marafivirus_PRO was mixed end-over-end for 1 h at 4 °C with GST-Bind resin (Millipore) that had been pre-equilibrated in lysis buffer. The lysate/resin slurry was poured into a gravity column and washed with ~20 column volumes of lysis buffer, followed by elution of the fusion protein with lysis buffer supplemented with 10 mM reduced glutathione (adjusted to pH 8.0 with NaOH). The GST tag was removed from each PRO domain (when applicable) using GST-tagged HRV 3C PreScission Protease, which was incubated with the eluted fusion protein in dialysis tubing overnight at 4 °C against 2 L of dialysis/gel filtration buffer (20 mM TRIS-HCl pH 8.0, 150 mM NaCl and 2 mM DTT). Tag-free PRO domains were separated from free GST and HRV 3C PreScission Protease using a Superdex75 (GE healthcare) gel filtration column. The concentration of each purified marafivirus PRO was quantified using a NanoDrop (ThermoFisher Scientific) instrument (A_{280}, ε/1000 BlVS 11.46, CSDaV 16.96, GSyV1 13.98, MRFV 8.48, OBDV 9.97, OLV3 13.98, TYMV 9.97 $M^{-1}cm^{-1}$).

MRFV_PRO_{C61A}-$HEL_{N\text{-}termDomain}$ and the WT version-containing cell pellets were resuspended in ice-cold lysis buffer (50 mM TRIS-HCl pH 8.0, 300 mM NaCl, 2 mM DTT and 5mM Imidazole) and lysed identically to the PRO domains. Cell lysates were clarified by

centrifugation (17,211 X g at 4 °C), and supernatants containing either protein were mixed end-over-end for 1 h at 4 °C with nickel-nitrilotriacetic acid resin (Qiagen) that had been pre-equilibrated in lysis buffer. The lysate/resin slurry was then poured into a gravity column and washed with ~20 column volumes of lysis buffer, followed by 10 column volumes of lysis buffer supplemented with 15 mM imidazole, followed by 10 column volumes of lysis buffer supplemented with 30 mM imidazole and finally eluted with lysis buffer supplemented with 250 mM imidazole. The eluted proteins were dialyzed against 2 L of dialysis/gel filtration buffer (20 mM TRIS-HCl pH 8.0, 150 mM NaCl and 2 mM DTT) overnight at 4 °C and then further purified using a Superdex75 gel filtration column. The concentrations of purified MRFV_PRO$_{C61A}$-HEL$_{N\text{-}termDomain}$ or WT were quantified using a NanoDrop instrument (A_{280}, ε/1000 = 39.42 $M^{-1}cm^{-1}$). UbVs were purified identically to MRFV PRO-HEL proteins.

Ubiquitin(1–75)-3-bromopropylamine (Ub-3Br) was prepared and purified as previously described (243, 244) for covalent coupling to MRFV/OBDV PRO. This is a version of human Ub (Uniprot P62987) lacking the terminal Gly. Purified Ub-3Br was dialyzed in 20 mM TRIS-HCl pH 8.0, 150 mM NaCl and 2 mM DTT, quantified by a Bradford protein assay and coupled with MRFV/OBDV PRO in a 2-fold molar excess at 4 °C for 16 h with the addition of Tris(2-carboxyethyl)phosphine hydrochloride to a final concentration of 5 mM. The resulting MRFV/OBDV PRO-Ub complexes were separated from excess Ub-3Br using a Superdex75 gel filtration column and quantified by A_{280} readings (ε/1000 MRFV_PRO-Ub 9.970 & OBDV_PRO-Ub 11.460 $M^{-1}cm^{-1}$).

3.2.4 Enzyme assays

All proteases (GST-fused or tag removed) were assayed against the fluorogenic substrate analogue 7-amino-4-methylcoumarin (AMC)–Ub (Ub-AMC; Boston Biochem) or the synthetic peptide analogue LRGG-AMC (GenScript). The latter substrate represents the C-terminal motif of ubiquitin as well as the linker between the RNA dependent RNA polymerase (RdRp) domain and major coat protein (CP) of the MRFV polyprotein. Reaction buffer for all assays consisted of 50 mM TRIS-HCl (pH 8.0), 150 mM NaCl and 2 mM DTT. Progress curves of fluorescence increase over time were collected on a Fluorolog-3 fluorimeter (Horiba Jobin Yvon) or a SpectraMax iD5 microplate reader (Molecular Devices). Reactions were 1 ml total volume in a quartz cuvette on the fluorimeter and 100 µl on the micro plate reader in a black, flat bottom 96-well microplate (Corning Life Sciences). In both situations, substrate was prepared first in reaction buffer to which enzyme was added immediately prior to readings. Monochromators were set to excitation of 345 nm and emission of 445 nm.

3.2.5 Polyubiquitin chain hydrolysis assays

200 ng of substrate (K48- or K63-linked poly-ubiquitin chains [Ub_3-Ub_6; Boston Biochem]) was incubated with 50 to 200 nM of each marafivirus DUB in a reaction buffer identical to the above-mentioned kinetics assay. Each reaction was incubated for 30 min at 25°C. Reactions were terminated with the addition of 2X SDS-PAGE loading buffer. Reactions were visualized by carrying out TRIS-Tricine PAGE (10 %) and subsequent detection using a Pierce silver stain kit (ThermoFisher Scientific).

3.2.6 Ubiquitin variant engineering and coupling

UbVs were generated based on a phage display library as described previously (245). The library was used against purified, immobilized GST-tagged proteases as previously described (245, 246). After selections, the most efficacious UbVs were chosen and cloned into pET53 expression vectors and were sent to our lab. pET53 vectors encoding the UbV gene were transformed into E. coli BL21(DE3) Gold. Protein expression was done identically to the proteases above and purification was done identically to the MRFV_PRO-$HEL_{N\text{-}termDomain}$ proteins, as the UbVs have an in-frame N-term 6xHis tag.

Once purified, UbVs were coupled with their respective proteases overnight at 4 °C with at least a 2-fold molar excess of UbV to force complex formation. The following day, the protein mixture was loaded onto a Superdex 75 gel filtration column and the protein complex was separated from excess UbV. Fractions pertaining to the protein complex were pooled and used for crystallization.

3.2.7 Protoplast transient gene expression assays

Arabidopsis thaliana (Columbia) leaves (50 to 60 leaves) were collected from 2–3-week-old plants. Mesophyll protoplasts were isolated as previously described (247, 248). Polyethylene glycol (PEG) -$CaCl_2$-mediated transfection was utilized with 0.5 x 10^5 to 0.5 x 10^6 protoplasts and microgram amounts of plasmid DNA. Protoplasts were harvested ~18-20 hours post-transfection and lysed as previously described (247). The soluble fraction of each reaction was isolated by centrifugation and was run on a 12 % or 4-20 % gradient SDS-PAGE gel (Bio-Rad). Proteins were transferred to a 0.45 μm nitrocellulose membrane using a semi-dry turbo transfer system (Bio-

Rad). The membrane was blocked as per the manufacturers recommendation and incubated with the appropriate primary antibodies at 4 °C overnight (Anti-V5-HRP (R961-25; ThermoFisher Scientific), Anti-c-Myc (9E10; Sigma Aldrich), Anti-FLAG (M2; Sigma Aldrich) or Anti-HA (12CA5; kindly provided by Dr. Peter Pelka). Where appropriate, the membrane was subsequently washed and then incubated with conjugated secondary antibody (goat anti-mouse IgG-HRP; Bio-Rad). Membranes were washed a final time and presented with enhanced chemiluminescent substrate (Luminata Forte; EMD Millipore) and detection was done using film (Amersham Hyperfilm or Pierce CL-Xposure film).

3.2.8 Tobacco (BY-2) cells propagation

Nicotiana tabacum (tobacco) BY-2 cells were purchased from Riken Research Center (experimental plant division) on semi solid medium. Immediately on arrival, cells were transferred to BY2 medium (Table 3.2 for components).

Table 3.2 BY-2 cell media components

Component	Concentration
Murashige and Skoog basal salt mixture	4.3 mg/ml
Sucrose	30 mg/ml
Thiamine hydrochloride	0.1 mg/ml
Myo-inositol	0.01 mg/ml
2,4-D sodium monohydrate	0.02 mg/ml
Potassium phosphate, dibasic, anyhydrous	0.2 mg/ml

*Adjust final pH to 5.8 with potassium hydroxide and autoclave

The cells took ~ 2 weeks to actively start dividing and develop their characteristic yellow colouration (constant conditions of: 28 °C, 130 rpm, dark). Once they became established, cells were subcultured 1:20 into fresh medium once every week aseptically.

3.2.9 Protein crystallization

MRFV PRO was crystalized using the vapour diffusion method at 15 mg/ml in a condition that contained 100 mM BIS-TRIS Propane (pH 7.5), 200 mM sodium acetate and 20 % PEG 3350. Crystals appeared after ~30 days at 4°C. Crystals of the MRFV PRO-Ub complex were also grown using the vapour diffusion method at 20 or 25 mg/ml in 100 mM phosphate citrate buffer (pH 3.8), 200 mM lithium sulfate and 25 % PEG 1000, which appeared after 1 day at 4°C. Crystals of the OBDV-Ub complex were grown using vapour diffusion method at 30 mg/ml in 20 % v/v jeffamine M-600 and 100 mM HEPES pH 7.5, which were found ~ 1 year after their initial setup. Immediately prior to X-ray data collection, single crystals of either PRO or Ub-bound PRO were swept through a cryoprotectant composed of the initial crystallization condition supplemented with 25 or 15% glycerol (PRO and Ub-bound PRO, respectively) and subsequently flash cooled in liquid nitrogen.

3.2.10 X-ray data collection and structure determination

X-ray diffraction data were collected in-house at 100 K using a Rigaku MicroMax HF X-ray generator and R-AXIS IV++ image plate detector. Data were indexed using XDS for MRFV PRO (249) and iMOSFLM for MRFV PRO-Ub (250). Scaling was done using Aimless (251) as a part of the CCP4 i2 program suite (252). For the unliganded MRFV PRO structure determination, molecular replacement (MR) was carried out using the crystal structure of TYMV PRO with its co-crystallized *E. coli* contaminant excluded (PDB code: 4A5U). MR was done using PHENIX.PHASER (253) and was followed by model building using PHENIX.AUTOBUILD (253). Iterative model building and refinement was done using COOT (254) and

PHENIX.REFINE (253). Structure determination for MRFV PRO bound to Ub was carried out almost identically; however, a multi component MR search was carried out using the structures of the un-liganded form of MRFV PRO (determined herein, waters removed) and ubiquitin (PDB code: 1UBQ). The preliminary structural data of OBDV PRO-Ub were determined identically to the MRV PRO-Ub structure, but the multi component MR search was done using the MRFV PRO molecule from the Ub-bound structure (Ub and waters removed) and ubiquitin (PDB code: 1UBQ). Crystallographic and refinement statistics are provided in Table 3.3 (section 3.2.5).

3.3 Results and discussion

3.3.1 Comparative sequence analysis of the PRO domains of six marafiviruses and one tymovirus

As shown in Fig. 3.2A the majority of the marafivirus proteome exists within one major polyprotein (~200 kDa) consisting of the core replication proteins needed in order for the virus to replicate its genome within the host (133, 146). Structural proteins are also present within the polyprotein as well as the sole endopeptidase PRO (133). The MT, PRO, HEL, RdRp and CP domains are consistently found in this arrangement within the polyprotein. The PRO domains are believed to be involved with the processing of the polyproteins into individual functional subunits by cleavage of the putative scissile bond(s) as observed for tymoviruses such as TYMV (161, 163, 164, 255).

An amino acid sequence alignment of the six marafivirus PRO domains we examined reveals considerable variability in the percent identity between the marafivirus enzymes and consistently low identity to TYMV PRO even though they appear to serve the same purpose (Fig. 3.2B & C). Regardless, the similarity between the marafivirus and TYMV PRO sequences was sufficient to identify marafivirus PRO domains based on alignments against the TYMV PRO sequence for which an X-ray crystal structure has been determined (Fig. 3.2B). The PRO domains from six marafiviruses were explored: *Blackberry virus S* (BlVS), *Citrus sudden death-associated virus* (CSDaV), *Grapevine Syrah virus 1* (GSyV1), *Maize rayado fino virus* (MRFV), *Oat blue dwarf virus* (OBDV) and *Olive latent virus 3* (OLV3). The DNA sequence for each protease was codon optimized for expression in *E. coli*. Amino acid sequences for the marafiviruses had been derived from genomic RNA sequences deposited in the National Center for Biotechnology

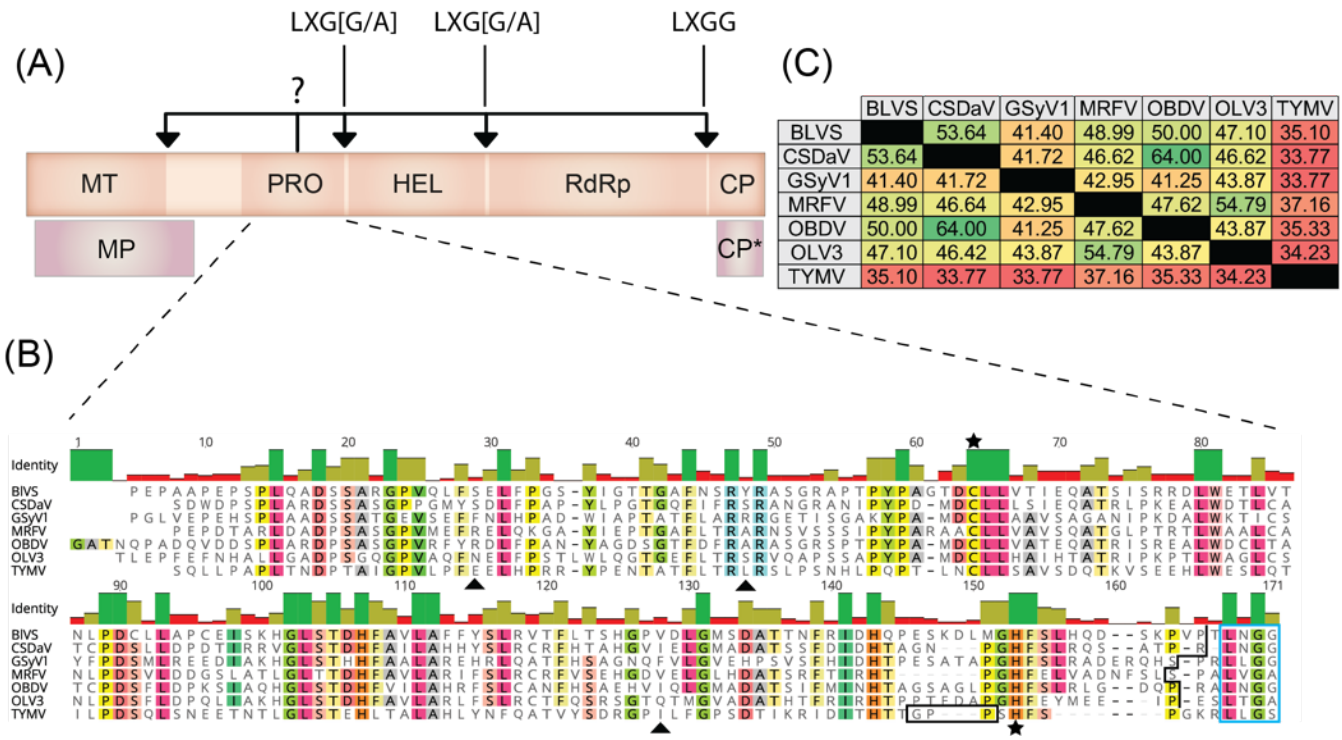


Figure 3.2 Marafivirus polyprotein arrangement and endopeptidase comparison.

(A) Simplified schematic of the typical arrangement of the sole polyprotein produced by marafiviruses and two additional proteins produced from genomic RNA. (MT) methyl transferase, (PRO) endopeptidase/deubiquitinase, (HEL) helicase, (RdRp) RNA-dependent RNA polymerase, (CP) major coat protein, (MP) movement protein, (CP*) minor coat protein. Arrows indicate junction points between each domain and the possible cleavage sites targeted by the endopeptidase. The four amino acid residues at each putative cleavage sites are indicated where "X" is any amino acid. *(B)* Sequence alignment of PRO domains from 6 marafiviruses and one tymovirus. *Blackberry virus S* (BlVS); *Citrus sudden death associated virus* (CSDaV); *Grapevine Syrah virus 1* (GSyV1); *Maize rayado fino virus* (MRFV); *Oat blue dwarf virus* (OBDV); *Olive latent virus 3* (OLV3); *Turnip yellow mosaic virus* (TYMV). Highlighted residues indicate agreement of ≥ 65 %. Stars indicate the catalytic cysteine and histidine of the active site and the triangle indicates a potential triad residue. Triangles indicate key residues. The box around the C-terminus GPP of TYMV is the mobile loop characteristic to tymoviruses. The stepwise line at the C-terminus of the marafivirus PRO domains indicate where the expression constructs used herein were terminated. Alignment was done in Geneious v.11.1.5. (http://www.geneious.com). *(C)* Percent identity matrix of all 6 marafivirus PRO domains and the PRO domain of TYMV. Values are presented in heat map format where green indicates a higher degree of sequence similarity. Multiple sequence align matrix values were determined using Clustal Omega (256).

Information. Fig. 3.2B shows the amino acid sequence of each PRO domain that was expressed. As determined here, functionally active marafivirus proteases are compact with an average sequence length and molecular weight of ~160 residues and ~17 kDa, respectively. The marafivirus PRO enzymes we studied share at least 40% sequence identity, with CSDaV and OBDV having a remarkable 64% sequence identity (Fig. 3.2C). Comparing the sole tymovirus protease (TYMV) with the sequences of the proteases from the marafiviruses consistently reveals, as expected, the highest degree of divergence. Interestingly, GSyV1 PRO has the least sequence similarity to any of the PRO domains of all marafivirus endopeptidases analyzed and shares the least similarity with TYMV (along with CSDaV) at 33.7%. In contrast, there is a 64% similarity between OBDV and CSDaV. These variations in sequence similarity can potentially be attributed to many factors from wide differences in hosts, host climate/ecosystem as well as viral vectors.

3.3.2 Marafivirus endopeptidases have auxiliary deubiquitinating activity

Protein ubiquitination is a highly conserved post-translational modification process that occurs in eukaryotes including plants, which regulates the function, trafficking and fate of protein substrates in the cell (16, 168). Ubiquitination involves the tethering of a Ub molecule(s) to a protein substrate by the combined efforts of Ub-activating enzymes, Ub-conjugating enzymes and Ub ligases (168, 257). Ub is typically conjugated to the substrate protein as a polyubiquitin chain (168). The first Ub molecule is covalently attached to a lysine on the surface of a substrate protein via the carboxy-terminal glycine of Ub, forming a covalent isopeptide bond with the lysine side chain ε-amino group (168, 257). Additional Ub molecules can be attached to the first conjugated Ub through linkages formed between one of seven lysine residues or Met1 on the surface of Ub (Lys6, Lys11, Lys27, Lys29, Lys33, Lys48 or Lys63) and the terminal Gly of the newly added Ub

molecule (16). Typically, Ub chains are of the same linkage type, but mixed poly-Ub chains are known to exist (168). The nature of the conjugation type and/or length of chain ultimately determines the function or fate of the protein substrate (257, 258).

In the model plant organism *A. thaliana*, 6% of all protein-encoding genes are purportedly linked to some facet of Ub modification (169) and 12 genes have been identified to date that code for functional Ub or Ub-like proteins such as RUB proteins (170), illustrating the large role that the ubiquitin system has in plants. Lys48 and Lys63 are the most well-understood poly-Ub chain types and the two most abundant forms in plants (259). Lys48 polyubiquitination (and Lys63 to a lesser extent) marks a substrate protein for degradation in plants by the ubiquitin proteasome system (UPS) (257, 260, 261), whereas Lys63 has many different roles in plant cells such as DNA replication/repair, iron homeostasis, endocytosis, nutrient transport, vacuolar sorting, protein synthesis, and immunity (170, 262).

Importantly, the process of ubiquitination is reversible, allowing Ub molecules conjugated to various substrate proteins to be uncoupled after the cellular function(s) dictated by ubiquitination is complete (16). This reversibility of the Ub system is carried out by cellular DUBs, of which there has been approximately 50 identified in *A. thaliana* alone (16, 170, 262, 263). Cellular DUBs are important in reversing Ub linkages to target proteins, but also are essential in processing Ub precursor proteins (170, 263). Not surprisingly, viruses have acquired the ability to exploit the Ub system to their advantage by encoding for multifunctional proteolytic enzymes (often cysteine proteases) that not only assist with viral replication by processing the viral polyprotein, but also acting as DUBs to shut down Ub-dependent host antiviral mechanisms (101, 149, 188, 264).

When it was discovered that TYMV PRO had bona fide DUB activity, it was suggested that additional plant viruses may possess this function as well, including marafiviruses (149, 153, 240). To assay for this potential DUB activity in marafiviruses, we chose to study the enzymatic activity of six PRO enzymes from a range of marafivirus species (Fig. 3.2B). Each PRO domain examined was based on the region known to exhibit endopeptidase and DUB activity from TMYV PRO and excluded the putative cleavage site at the putative PRO|HEL junction of each viral polyprotein as shown in Fig. 3.2B. Each construct contained the conserved Cys and His residues (Fig. 3.2B, denoted with stars) that form the papain-like cystine protease active site, forming the catalytic dyad in which a cysteine nucleophile and histidine base work in concert to hydrolyze the scissile bond (16, 264). The recombinant PRO domains were quite stable, and each could be purified to homogeneity (Fig. 3.3A) The fluorogenic substrate Ub-7-amido-4-methylcoumarin (Ub-AMC) was used to assess the DUB activity for each enzyme (265). The fluorogenic substrate on its own has a relatively low signal; however, hydrolysis by DUB enzymes liberates the AMC fluorophore from Ub, which de-quenches the fluorophore and results in a measurable signal (265). As seen in Fig. 3.3B, the PRO domains of BlVS, CSDaV, GSyV1, OBDV and OLV3 (black, red, purple blue, and green, respectively) were incubated with Ub-AMC at a constant concentration and fluorescence measurements were taken over a 20 min period at ambient temperature. A clear increase in fluorescence was observed for all five of the viral enzymes over time compared a control lacking enzyme, which showed no appreciable Ub-AMC hydrolysis. These data confirm that marafivirus PRO enzymes do exhibit DUB activity, further expanding the number of known viruses known to encode this activity. As all of the structural data presented here was for the marafivirus type member PRO domain (MRFV PRO), a more thorough progress curve was

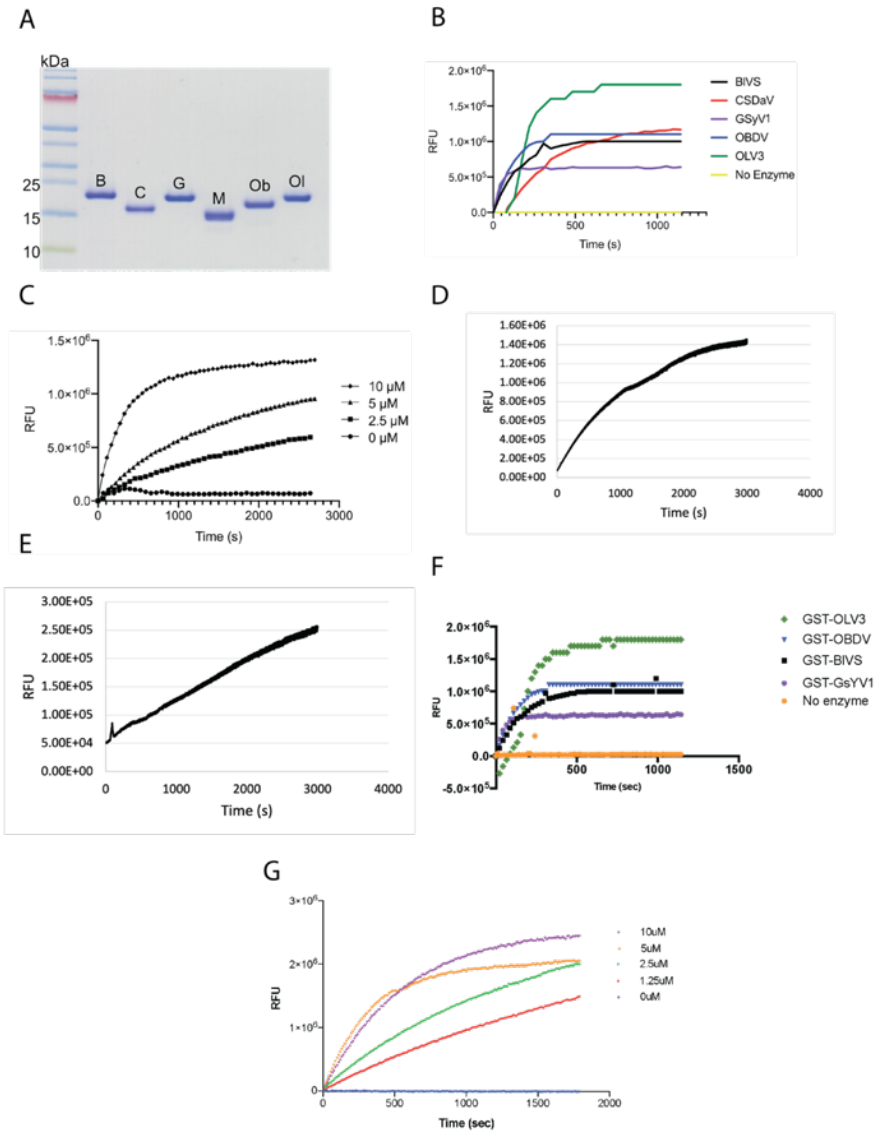


Figure 3.3 Purified proteases and Ub-AMC progress curves.

A) 10% TRIS Tricine SDS-PAGE gel of purified PRO domains of 6 marafiviruses (*Blackberry virus S* (B); *Citrus sudden death associated virus* (C); *Grapevine Syrah virus 1* (G); *Maize rayado fino virus* (M); *Oat blue dwarf virus* (Ob); *Olive latent virus 3* (Ol)). B) Fluorescence vs. time progress curve of marafivirus proteases with Ub-AMC. Substrate was used at a final concentration of 200 nM and enzyme at 5 µM. Curves are coloured with respect to enzyme (Black-*Blackberry virus S* (BlVS); Red-*Citrus sudden death associated virus* (CSDaV); Purple-*Grapevine Syrah virus 1* (GSyV1); Blue- *Oat blue dwarf virus* (OBDV); Green-*Olive latent virus 3* (OLV3); Yellow-substrate only). C) Fluorescence vs. time progress curve of MRFV PRO with Ub-AMC. Substrate was used at a final concentration of 200 nM and enzyme concentration was varied as indicated. D-F) Fluorescence vs. time progress curve of GST-PRO fusion proteins with Ub-AMC (D, CSDav; E, MRFV; F, as indicated; G, TYMV). Enzyme and substrate concentrations were 5 µM (D-F, G varied as indicated) and 200 nM, respectively.

generated to show enzyme concentration dependence (Fig. 3.3C). It can be seen that there is a clear concentration dependent rate of increase in fluorescent signal directly proportional to the relative amount of enzyme present per reaction. These findings confirm that marafivirus PRO domains are also DUB enzymes.

It was also important to verify that these enzymes were active as GST-fusion proteins, as the UbV phage display library relies on the DUBs to be immobilized by GST-specific plates upon screening using the library. It was imperative to determine that even with GST bound to the N-term of the proteases, that they could hydrolyze Ub-AMC, indicating that GST does not occlude the active site or distort how the active site forms. As seen in Fig. 3.3D-G, the proteases still have activity towards the substrate, indicating that the active sites (the targets of the library) should still be accessible and able to be effectively probed. GST-TYMV PRO (3.3G) was also used to test activity against Ub-AMC and for subsequent generation of TYMV PRO-specific UbVs, which is now a central focus of a colleague's thesis work.

As above mentioned, the cellular outcome of ubiquitination is often dictated by the topology of the Ub scaffold tethered to the target protein in the form of poly-Ub chains (16). To evaluate the substrate recognition ability of these proteases towards poly-Ub chains, each enzyme was co-incubated with either K^{48}/K^{63} poly-Ub chains (3 to 6 Ub molecules in length) to determine their substrate specificity. Upon mixing enzyme with each Ub chain type, the presence of di/mono-Ub should accumulate, and higher molecular weight species diminish if Ub chain hydrolysis is occurring. As seen in Fig. 3.4A, B, C & F, BlVS, CSDaV, GSyV1 and OLV3 PRO all appear to hydrolyze both K^{48}/K^{63} poly-Ub chains. A clear accumulation of di/mono-Ub is present for both substrate types and increases with higher enzyme concentration. Interestingly, MRFV and OBDV PRO (Fig. 3.4D&E) only appear to act on K^{48} poly-Ub chains. When comparisons are drawn

between the proteases at the sequence level, MRFV PRO is more similar in sequence to BlVS and OLV3 PRO (~49 and 55 %, respectively—Fig. 3.2C) than in comparison to OBDV PRO (~48 %), whereas OBDV PRO is more similar to BlVS PRO (50 %) and CSDaV PRO (64 %); however, the preferences of MRFV and OBDV PRO towards Ub-substrate is shared. The preference of MRFV and OBDV PRO for K^{48} poly-Ub chains illustrates a surprising difference between these marafivirus endopeptidases. It is possible that BlVS, CSDaV, GSyV1 and OLV3 PRO share a common structural feature that allows for broader substate specificity and is absent from MRFV and OBDV PRO. Unfortunately, there are too many differences between the enzymes at the primary sequence level to identify a region or motif that could be responsible for the difference in substrate specificity and 3D structural analysis would be needed to gain further insight. For comparison, TYMV PRO is able to hydrolyze both K^{48} and K^{63} poly-Ub chains, which is shared with BlVS, CSDaV, GSyV1 and OLV3 PRO (149). This suggests the presence of nuanced, shared similarities/differences exist between these marafivirus and tymovirus DUBs that impart specificity towards their substrate(s) even though, from a primary sequence point of view, no definitive motifs underlying these functional differences are apparent.

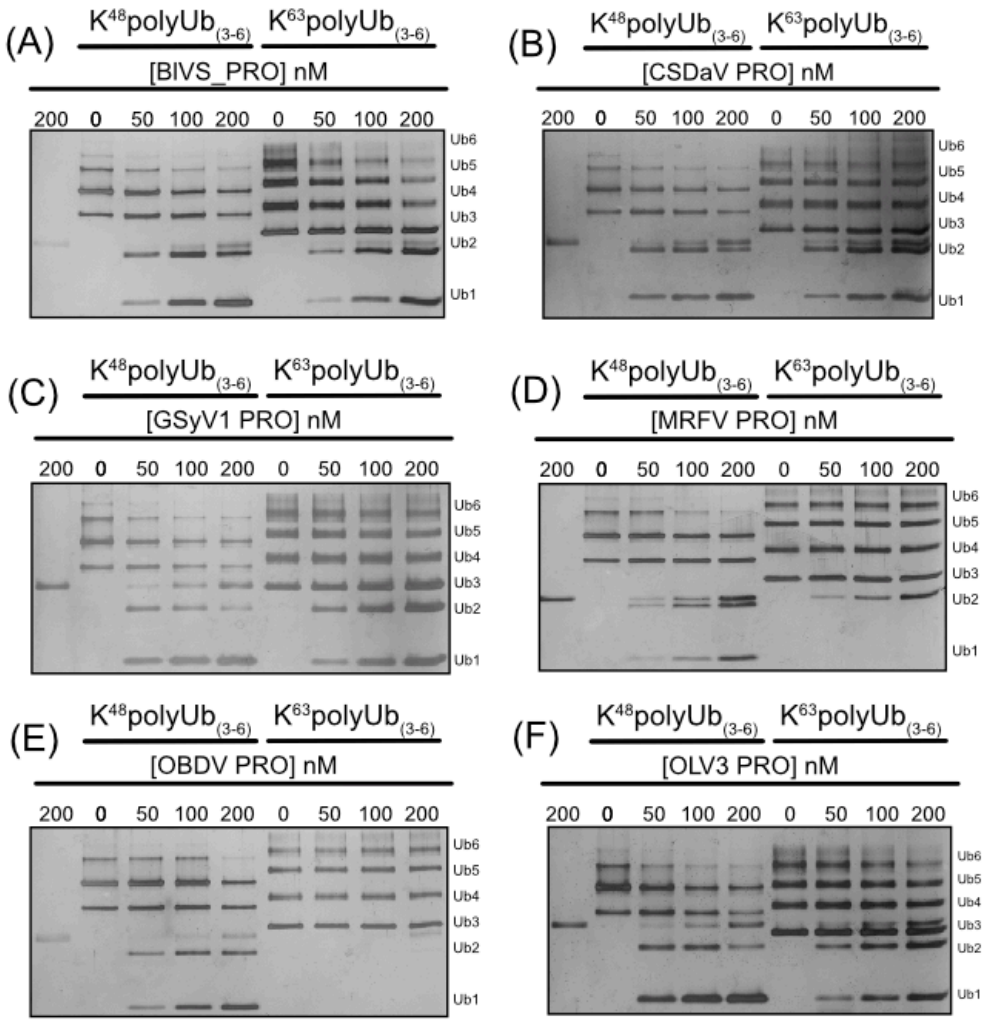


Figure 3.4 Qualitative analysis of poly-Ub chain hydrolysis by marafivirus proteases.

K^{48}/K^{63}poly-Ub chains in the presence or absence of variable concentrations of the indicated marafivirus proteases/DUBs. Lane 1 for each assay contains protease alone as a reference. Ub chain lengths following hydrolysis are indicated.

3.3.3 CSDaV PRO A. thaliana protoplast assays

We commenced our study of marafivirus proteases using CSDaV PRO as our prototype enzyme. It was an appealing target as the virus infects orange trees grafted onto the rootstocks of the Rangpur lime (143). Interestingly, only the rootstock is infected by CSDaV (143). Being that the long-term goal of this project was to make a transgenic plant resilient to viral infection, it was attractive to potentially make a transgenic rootstock that would yield fruit that was not genetically modified, as WT orange trees would simply be grafted onto the transgenic rootstock. Unfortunately, CSDaV PRO ultimately proved challenging in functional and structural characterization as it was recalcitrant to crystalizing and complex formation, and we subsequently moved to studying MRFV PRO to gain structural insights into marafivirus proteases. Nevertheless, we were able to carry out several transient expression assays of CSDaV PRO in *A. thaliana* mesophyll protoplasts, which will be outlined in this section.

The purpose of these studies was to establish a means to study how the UbVs generated against our DUBs would stifle polyprotein processing in plant cells using *A. thaliana* protoplasts (Fig. 3.5A), which can be presented with plasmid DNA for transient gene expression assays. The general idea of these experiments was to initially transfect cells with DNA coding for CSDaV PRO, along with DNA coding for a section of the polyprotein that would likely be cleaved by PRO (i.e. PRO|HEL). This would have told us if the enzyme indeed does cleave at the PRO|HEL junction. We would have then been able to introduce plasmid DNA coding for the UbV and monitor how proteolysis of the fusion protein would be inhibited by the UbV. Many different constructs were prepared pertaining to CSDaV proteins. As seen in Fig 3.5B, the very first construct we prepared was N-term V5-tagged CSDaV PRO (~17.5 kDa), which we were able to see being produced in *A. thaliana* protoplasts.

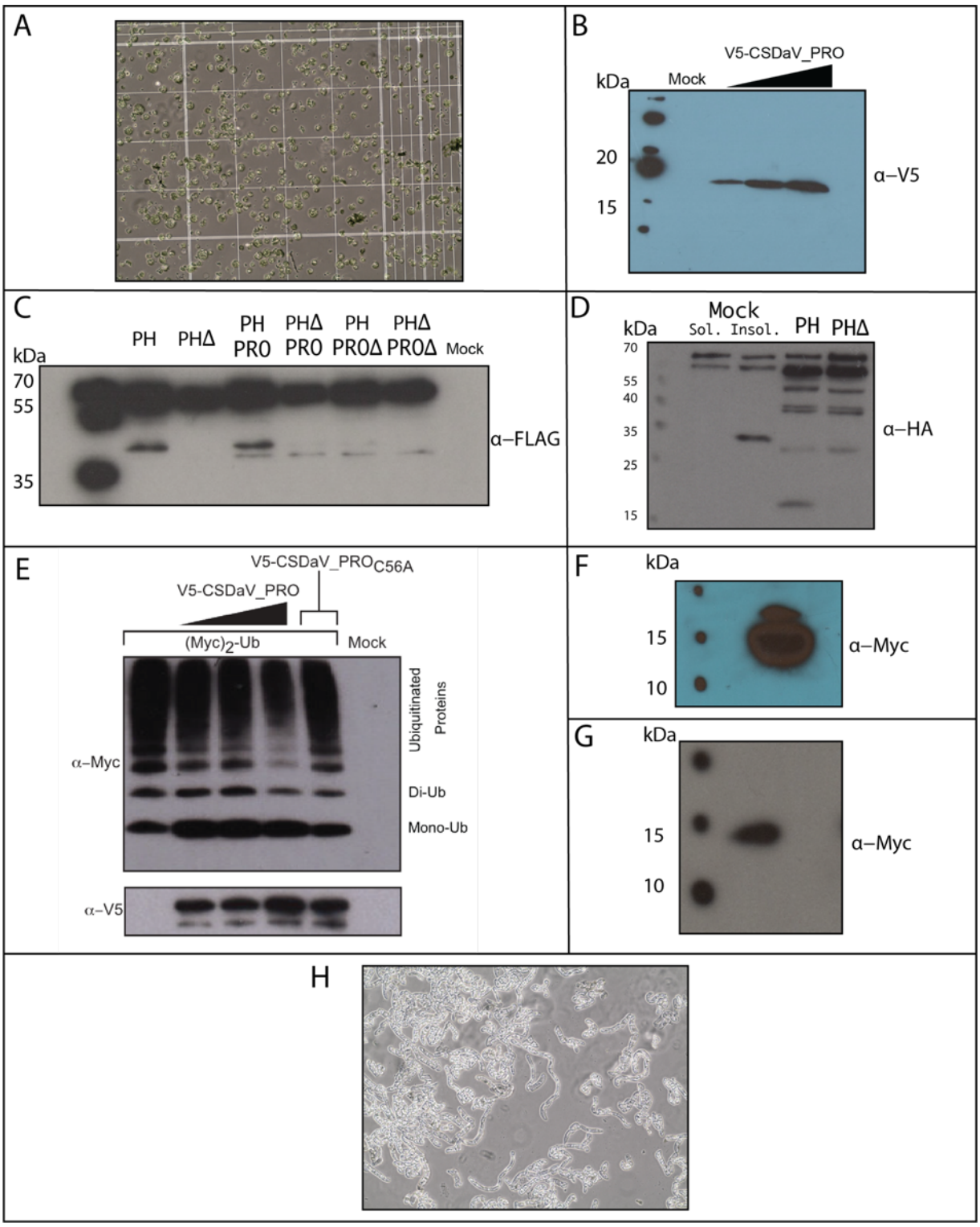


Figure 3.5 *A. thaliana* transient gene expression assays.

A) Purified *A. thaliana* mesophyll protoplasts on a hemocytometer. B) Transient gene expression of CSDaV PRO with increasing amounts of pHBT95 plasmid transfected (5, 10 and 20 µg).

Detection done using anti-V5 1°Ab. Mock lane was protoplasts isolated identically, however, no plasmid was delivered to cells. C) All plasmids from pHBT95 and contained various inserts (PRO|HEL, PRO|HEL C56A, PRO, PRO C56A). Detection done using anti-FLAG 1°Ab. D) Similar transfection to C) using anti-HA 1°Ab. E) Ubiquitinated host cell proteins in the presence or absence of catalytically active/inactive V5-tagged CSDaV PRO. Smears and bands in the top panel blot are indicative of soluble host cell proteins that have been conjugated with $(Myc)_2$-Ub. The bottom panel contains identical samples blotted using an anti-V5 antibody probing for the transiently expressed WT or C56A protease. F & G) Transient gene expression assay of pΩ-$Myc_2(Ub^{75}AA^{76})$ or pΩ-Myc_2(CSDaV UbV), respectively and blotted using anti-Myc 1°Ab. H) Isolated *N. tabacum* BY-2 cells taken from suspension culture.

We next designed a construct that encoded for a purported cleavage target, encoding for the entirety of the PRO and HEL domains with an N-term HA tag and C-term FLAG tag. The full-length protein has a molecular mass of 59.2 kDa, which we were able to detect by immunodetection (Fig. 3.5C) in soluble cell lysates of transfected protoplasts that had been allowed to incubate overnight for protein production. Interestingly, it appeared that the enzyme was cleaving *in cis* at its own C-term (lane 2), as we were able to detect a smaller molecular mass protein which would have been the FLAG-tagged HEL domain (41.4 kDa), which we did not see for the catalytic knockout construct with a Cys56Ala mutation (lane 3). Interestingly there was no increase in the cleavage product when a construct coding for WT PRO was also delivered, showing no discernible cleavage *in trans* (lane 4). Very little cleavage product was detectable in the catalytic knockout fusion protein transfected with WT PRO (lane 5), as we believe the enzyme potentially has higher affinity for itself *in cis*. Oddly, there is very little cleavage product in the WT PRO|HEL fusion protein transfected with Cys56Ala PRO (lane 6), which slightly confounds the preference for *in cis* cleavage, however, there is clearly no cleavage product when both mutant plasmids are transfected (lane 7). The results from these experiments were promising as we had the notion that we potentially only had to deliver one plasmid for monitoring cleavage and not a double transfection that would include the cleavage target and the enzyme separately, which would have made for a more streamlined and manageable workflow as *A. thaliana* protoplast isolation and transient gene expression assays proved to be somewhat incosistent and challenging.

A similar experiment was performed to see if we were able to detect the N-term cleavage product (HA-tagged PRO), however the antibody seemed to have a lot of cross reactivity towards cellular proteins (Fig. 3.5D). There was an intense band resembling the mass of the fusion protein that runs, in mass, almost identically to a cellular protein (darkest band ~59 kDa, lanes 4 & 5),

although it is hard to confirm definitively if that was the protein of interest. Strangely, there were multiple bands in the transfected cells (WT or C56A) that don't correspond to cleavage products with molecular masses in between the fusion protein and the cleavage product (lanes 4 & 5). We were able to see a band that agreed with the molecular pass of HA-tagged PRO (17.5 kDa) in the WT fusion protein sample (lane 4), which was not present in the catalytic knockout (lane 5), however the amount of nonspecific binding of the antibody to cellular proteins made this hard to confirm. Regardless, the FLAG antibody seemed to be working much better, so we decided to settle on that means of cleavage detection.

We were able to secure a plasmid compatible with *A. thaliana* transient gene expression of Ub with a double N-term Myc tag (section 3.1.2.2). We first wanted to assess global DUB activity of a CSDav PRO towards ubiquitinated plant cell proteins. Several viral DUBs have been shown to act non-specifically and can deconjugate Ub molecules from a wide array of ubiquitinated proteins in mammalian cells (101, 188, 266–268). TYMV PRO is not reported to act as a global DUB towards Ub-conjugated cellular *A. thaliana* proteins when expressed in cells as a full-length protein linked to the methyl transferase domain (149); however, our findings suggest that CSDaV PRO is able to remove Ub conjugates from a range of cellular proteins of different masses (Fig. 3.5E). The expression plasmid (pΩ) coding for c-Myc-tagged Ub was used to transfect purified protoplasts with and without the addition of V5-tagged CSDaV PRO (or the catalytic knockout C56A) with increasing concentrations of plasmid encoding for the active enzyme. Following an incubation time of ~18 h, protoplasts were harvested and lysed. Immunodetection was carried out on soluble cell lysates using an anti-c-Myc antibody to probe for ubiquitinated proteins originating solely from the transient expression of the introduced Ub-coding plasmid (Fig. 3.5E), as indicated by a dark smear pattern. Upon introduction of plasmid coding for V5-tagged CSDaV PRO (Fig.

3.5E, lanes 2-4), the intensity of the smearing pattern decreases as the amount of expression plasmid increases, accompanied by an accumulation of mono-Ub (smallest band). This result suggests that CSDaV PRO is capable of causing a global reduction in protein ubiquitination, decoupling Ub from a variety of Ub-conjugated cellular substrates of various masses. No apparent difference is seen between protoplasts only transfected with c-Myc-Ub-coding plasmid or doubly transfected with c-Myc-Ub-coding plasmid with plasmid coding for catalytically inactive CSDaV PRO (Fig. 3.5E, lane 5).

It has been shown that TYMV PRO deubiquitinates the RdRp of the TYMV replicase, rescuing the polymerase from proteasomal degradation via K48-polyubiquitination (149, 167). The DUB activity of TYMV PRO thus decreases RdRp turnover and appears to contribute to viral infectivity in plant cells (149, 167). Interestingly, TYMV PRO is also able to hydrolyze K63 poly-Ub chains (149), but its intended target substrate in the cell is unknown. The role of K63 uncoupling by viral plant DUBs is not well understood compared to mammalian-system affecting viral DUBs, which degrade K63 linkages to suppress innate immune signalling cascades (188, 244, 264, 269). It could be that TYMV PRO has unintended cross-reactivity towards K63 poly-Ub chains even though its main target is K48 linkages on the RdRp; however, TYMV PRO shows some degree of specificity as it opts to not aberrantly uncouple Ub linkages from nonspecific host proteins (149). The opposite may be true for CSDaV. It has marked activity against K63 poly-Ub chain in addition to K48 poly-Ub and exhibits DUB activity towards a wide array of cellular proteins. BlVS, GSyV1 and OLV3 also exhibit activity against K48 and K63-poly-Ub (Fig. 3.4) and may also exhibit broad specificity against ubiquitinated cellular substrates. It is possible that these enzymes uncouple Ub from a variety of cellular targets to promote viral replication, not just the viral RdRP. This is true for mammalian +RNA viruses that encode DUB enzymes (264) and

additional studies may reveal this to also be the case for plant +RNA viruses. In contrast, MRFV and OBDV PRO show a finer degree of substrate recognition compared to the other viral PRO enzymes, including TYMV PRO, showing activity only against K48 linkages, suggesting preference towards preventing RdRp degradation by the 26S proteosome or other aspects of K48 poly-ubiquitination in plants, such as plant development, hormone signaling and cell cycle mediation (171).

Using pΩ as a template plasmid, we were able to generate a construct coding for a mutant version of Ub with the terminal two Gly (^{75}GG^{76}) residues mutated to Ala residues. This version of ubiquitin cannot be conjugated to cellular substrates as linkage occurs between a cellular protein's Lys residue forming an isopeptide bond with the terminal Gly carboxy group of Ub. Its purpose was solely to act as a control; however, we were able to construct it and detect it in our protoplast system (Fig 3.5F). We were also able to generate a plasmid encoding for the UbV, which was also readily detectable (Fig. 3.5G). We were now poised to attempt to see how introducing the UbV affects polyprotein processing. Unfortunately, our *A. thaliana* protoplast system completely failed to function at this point. We were not able to regularly produce robust protoplasts that would survive the isolation process and when we were able to make it to the transfection step, no transient gene expression was detectable. Even prior, readily detectable proteins (V5-PRO or HA-PROHEL-FLAG) were not yielding any signal and the system, reagents, plant propagation techniques was troubleshooted for months prior to abandoning it.

We were able to source *Nicotiana tabacum* cells (BY-2) from Japan (Riken Plant Science Centre) that grew in suspension indefinitely to very high densities with proper subculturing and aseptic technique. This idea was promising as transient gene expression assays have been reported in this cell line, many cells could be obtained with very little effort, and our current plasmid arsenal

was compatible in *N. tabacum* (270). We were able to establish the cell line in our lab (Fig 3.5H), however the cells died during a power outage, and we had not been able to make cryogenic stocks of them to re-establish the line. At the same time, structural characterization and binding tendencies of the UbV generated against CSDaV PRO were proving challenging and not as promising as identical studies with MRFV PRO and its UbV, so we switched our attention to MRFV PRO.

3.3.4 MRFV PRO facilitates polyprotein processing

Autocatalytic viral polyprotein processing by a protease domain(s) encoded within the polyprotein represents a remarkably efficient mechanism of protein expression by +RNA viruses (87, 163, 271–273). TYMV's polyprotein has been shown to have at least two cleavage sites, which is carried out by the PRO domain (148, 163, 274). They are located between the PRO|HEL and HEL|RdRp junctions. As above mentioned, MRFV PRO was used in this and all subsequent experiments as it was better behaved *in vitro* for biochemical and structural studies, as well as having greater North American agricultural relevance. Furthermore, being the type member of the genus, we believe that it is most representative. To begin exploring the polyprotein processing by the MRFV PRO domain, two *E. coli* protein expression constructs were designed for the recombinant production of a subsection of the MRFV polyprotein spanning the PRO-HEL region. Two versions of the region were generated, one that contained a catalytically active PRO domain (WT) and another where the active site cysteine of PRO had been mutated (C61A). For both proteins, the PRO domain was expressed in its entirety, whereas only the N-terminal domain (ATP-binding domain) of the helicase was included. Importantly however, the region contained the putative LVGA recognition site at the PRO|HEL junction (Fig. 3.2A & B; cyan box*)*. We predicted the site would be cleaved by the PRO domain even though it contained a GA motif at

the C-terminus, which is atypical for a DUB, which usually cleaves after a di-glycine motif (GG). The complete PRO-HEL fusion would have been ~60 kDa, while the truncated form is ~44 kDa, which proved amenable for expression in *E. coli*. Cleavage at the predicted site by PRO would result in a ~19 kDa N-terminal, His_{10}-tagged PRO (with an enterokinase site in between the affinity tag and PRO), as well as an untagged version of the ATP-binding domain of the MRFV helicase (~25 kDa) that would not be captured through affinity purification if cleavage by PRO were to occur.

Fig. 3.6A depicts the gel filtration chromatograms of the WT and C61A PRO-$HEL_{N\text{-}TermDomain}$ proteins. As shown for the active site mutant (dashed line), only one significant species is present at an elution volume of ~65 ml, whereas in the WT trace (solid line), two species are present, suggesting autocatalytic cleavage of the PRO-$HEL_{N\text{-}TermDomain}$ by PRO. For the WT protein, the larger species shares a nearly identical elution volume as the species containing the PRO active site mutant and likely represented the intact PRO-$HEL_{N\text{-}TermDomain}$. Furthermore, for the WT protein, a second, lower molecular weight species appears at an elution volume of ~ 80 ml. As stated above, cleavage at the proposed LVGA PRO|HEL junction would generate two additional proteins, only one of which (PRO) would be retained on a nickel affinity column as it is the only one that retains a His_{10} affinity tag. Fig. 3.6B shows an SDS-PAGE gel of the species obtained from the WT gel filtration experiment. The intact full-length protein would have a theoretical molecular weight of ~ 44 kDa, whereas the proteins generated by PRO mediated cleavage would be ~19 kDa (PRO domain) and ~25 kDa (HEL ATPase domain). The processed helicase portion of the WT protein was lost in the purification process. Fig. 3.6A & B demonstrate that the WT

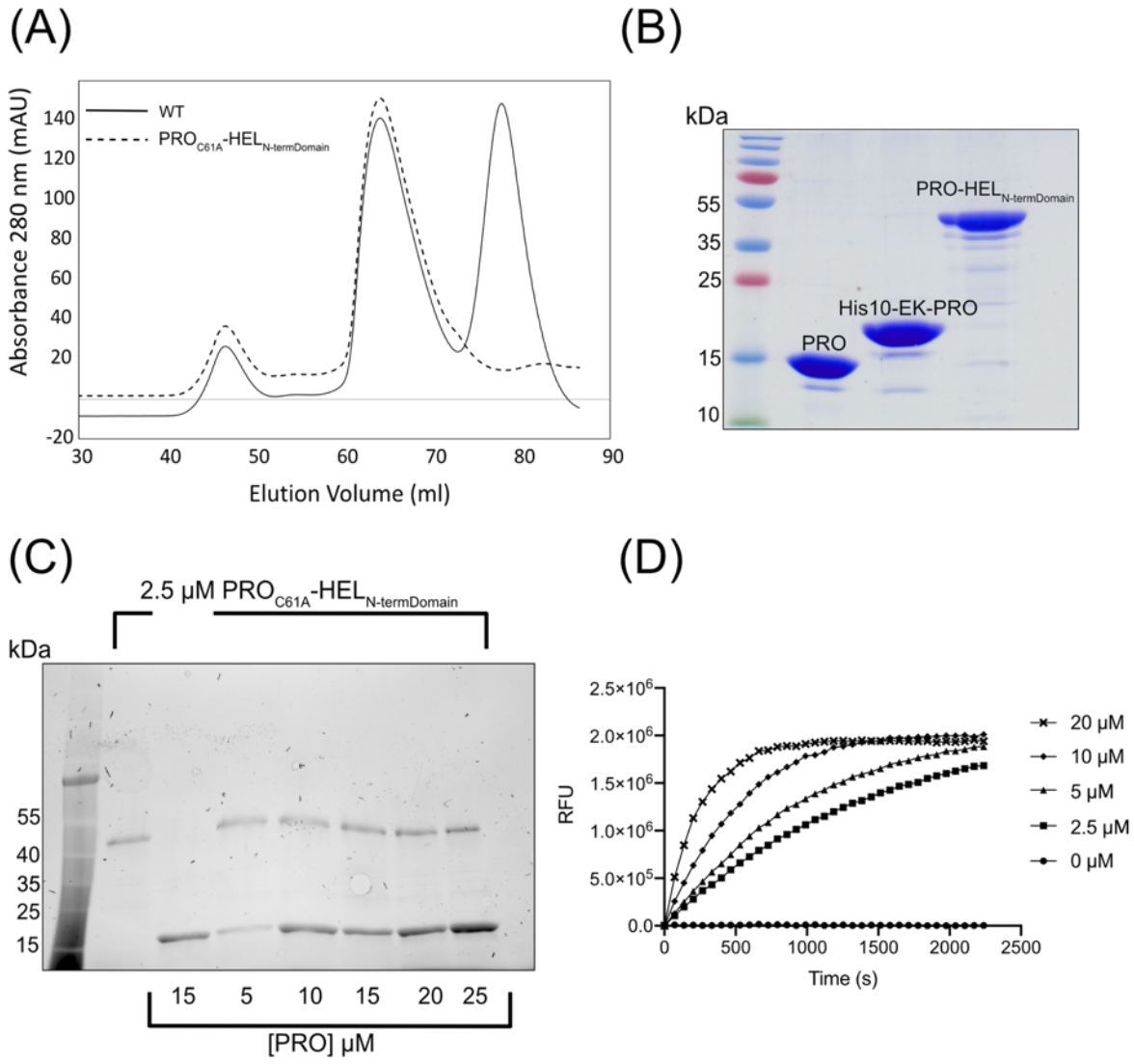


Figure 3.6 MRFV PRO cleaves at the PRO|HEL junction *in cis* and has catalytic specificity towards the C-terminal residues of ubiquitin.

(A) Gel filtration chromatograms of the purification of MRFV PRO$_{C61A}$-HEL$_{N\text{-}TermDomain}$ and WT. *(B)* 12 % Coomassie-stained SDS PAGE gel of MRFV PRO from pGEX-6P-1 (~16.4 kDa), in *cis* cleaved MRFV PRO from WT MRFV PRO-HEL$_{N\text{-}TermDomain}$ (~19.4 kDa) and uncleaved WT MRFV PRO-HEL$_{N\text{-}TermDomain}$ (~44.1 kDa). *(C)* 2.5 µM of purified MRFV PRO$_{C61A}$-HEL$_{N\text{-}TermDomain}$ was incubated with increasing concentrations of MRFV PRO for 30 min at 25°C. Each reaction was subsequently loaded onto a 12 % stain-free SDS-PAGE gel (Bio-Rad). *(D)* Fluorescence vs. time progress curve of MRFV PRO with LRGG-AMC. The substrate was used at a concentration of 25 µM and enzyme concentration was varied as indicated. Kinetic plots were designed in GraphPad Prism v.8.4.3.

PRO-HEL$_{N\text{-}TermDomain}$ is processed at the predicted junction; however, not to completion, which was also seen with CSDaV PRO|HEL in *A. thaliana* protoplasts (Fig. 3.5C). The identities of both MRFV WT species were confirmed with MS/MS.

To assess the nature of how PRO is acting to process the PRO|HEL junction, be it in *cis* and/or *trans*, a cleavage assay was carried out using the C61A mutant. As seen in Fig. 3.6C, a constant concentration of the C61A mutant of PRO-HEL$_{N\text{-}TermDomain}$ was incubated with an increasing concentration of active MRFV PRO domain. The PRO domain used in the experiment was identical to that used in the ubiquitin hydrolase assays (Figs. 3.3 & 3.4). The putative LVGA cleavage site in the C61A mutant of PRO-HEL$_{N\text{-}TermDomain}$ was not altered and thus remained cleavable. The results from Fig. 3.6C indicate that even in a 10-fold molar excess of catalytically active PRO, the C61A mutant fusion protein was not processed in *trans* by the PRO domain; no species smaller than the full-length protein, excluding the added PRO domain (~16 kDa), are visible. Should processing have occurred, both the larger PRO portion (~19 kDa) of the PRO-HEL$_{N\text{-}TermDomain}$ and ATP-binding domain of the helicase (~25 kDa) would have appeared. Taken together, these results indicate that the marafivirus PRO domain extracts itself from the viral polyprotein by in *cis* cleavage of the PRO|HEL junction. These findings are similar to what we saw in *A. thaliana* protoplasts with CSDaV PRO, which also preferred *in cis* cleavage (Fig 3.5C). This differs from what is observed in the closely related tymoviruses in which cleavage at the PRO|HEL junction appears to occur both in *cis* and *trans* (163). However, exclusive in *cis* cleavage is not uncommon in plant-affecting +ssRNA viruses, as is seen in members of the families *Potyviridae* and *Closertoviridae*, whose endopeptidases have been shown to also act solely in *cis* (60, 61). It could be that the release of PRO from the HEL domain enables PRO to adopt a fold

that allows for in *trans* cleavage at the HEL|RdRp and RdRp|CP junctions, but this would require further studies.

As above mentioned, many proteases from +ssRNA viruses have auxiliary functions that aid in viral replication, including DUB activity. The C-terminus of Ub is composed of the four amino acid motif LRGG. As demonstrated above and consistent with a number of other viral DUBs, marafivirus DUBs can recognize the C-terminus (LRGG) motif of Ub and cleave the bond downstream of the di-glycine motif (16). Interestingly, the sequences at the PRO|HEL, HEL|RdRp and RdRp|CP junctions all mimic the C-terminus of Ub (Fig. 3.2A). Indeed, the predicted RdRp|CP junction in MRFV has an exact LRGG sequence and very likely a PRO cleavage site. Though we did not carry out a cleavage assay as detailed as the PRO|HEL assay described above, Fig. 3.6D illustrates the ability of MRFV PRO to hydrolyze the synthetic fluorogenic peptide LRGG-AMC at the scissile bond between the terminal Gly and AMC. A clear concentration dependent trend is seen with increasing amounts of enzyme over time. In light of the X-ray structure of MRFV PRO described below, it is not surprising that LRGG-AMC is a poor substrate compared to Ub-AMC, as the peptide would have minimal interactions with the enzyme compared to Ub. Nevertheless, the LRGG-AMC assay indicates that marafivirus PRO domains are able to recognize LRGG alone and very likely cleave at the RdRp|CP junction to liberate the major CP from the replicase proteins. Given our cleavage data, the HEL|RdRp junction is also most likely processed by PRO. The Ub-like LXG[G/A] sequence at the HEL|RdRp junction is conserved amongst the marafiviral polyproteins (Fig. 3.2A) and previous results of tymoviruses processing at this junction have been demonstrated (148, 255, 274). Whether HEL|RdRp and RdRp|CP of the marafivirus polyprotein are also processed exclusively in *cis* and the temporal nature of these events remains to be determined.

3.3.5 Crystal structure of MRFV PRO

Initially, CSDaV PRO was attempted for structural studies, but was recalcitrant in crystallization. MRFV PRO proved to be more amenable to crystallization, with the enzyme crystalizing in space group $P2_1$ and its three-dimensional structure being determined to a resolution of 1.9 Å using the TYMV PRO X-ray structure (PDB code: 4A5U) as a molecular replacement search model. The protease adopts a compact, three-domain fold (Fig. 3.7A), although its C-term conformation is held in place by a neighbouring copy of the enzymes (Fig. 3.7 D& E). As seen in Fig. 3.7A, the first domain (blue panel) is composed of a β-hairpin (β1↑ β2↓) and α-helices α1 and α2. The second domain (green panel) is predominantly a well ordered five-helix bundle (α3-α7). Helix α3 contains the catalytic cysteine nucleophile (C61) with its solvent exposed thiol group at the N-terminal end of α3 (Figs. 3.7 A(arrows) & B). The third domain (orange panel) is comprised of a three-stranded β-sheet (β4↑ β3↓ β5↓) whose curved, open face packs against helices α3 and α7 of domain II. Domain III terminates with what appears to be a flexible loop that is solvent exposed until is inserts itself into the active site of a neighbouring copy of the enzyme. This loop contains the active site histidine (H144), which is ~17 Å from the catalytic cysteine. Based on sequence alignments and a successful molecular replacement experiment, MRFV PRO is believed to have a simple Cys/His catalytic dyad similar to TYMV PRO (16).

The ~17 Å distance between the side chains of C61 and H144 does not comprise a functional active site. The asymmetric unit (ASU) is composed of two copies of MRFV PRO (Fig. 3.7C) with most interactions occurring between the alpha helical bundles of domain II of each molecule. As seen in Fig. 3.7D, the C-terminal tail of each copy of protein in the ASU traverses into

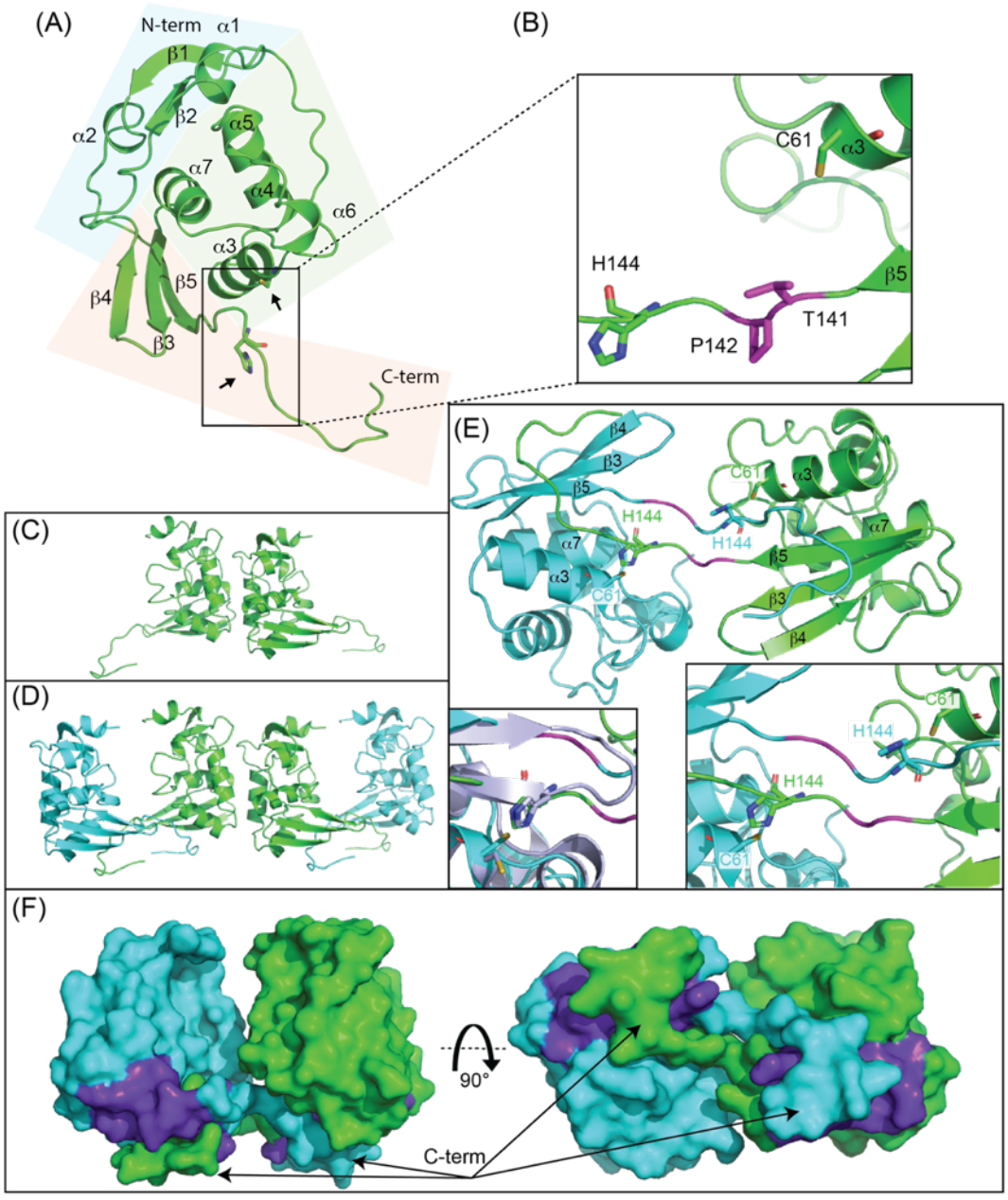


Figure 3.7 Crystal structure of MRFV PRO.

(A) Overall structure of MRFV PRO determined at 1.9 Å resolution. Individual domains of PRO are shown in blocked colours (domains 1, 2 and 3; blue, green and orange, respectively). The catalytic residues are indicated with arrows. *(B)* Close-up of the active site architecture of MRFV PRO. *(C)* Asymmetric unit of MRFV PRO crystals. *(D)* Interaction of MRFV PRO symmetry mates. *(E)* Closeup on interaction of MRFV PRO symmetry mates, one of each are shown in green and cyan. Labeling is identical to (*A)* & *(B)*. Inset is a closeup of the symmetry mate active site with TYMV superposed in lavender (left) and a closeup on the active site of just MRFV PRO symmetry mates (right) *(F)* Surface representation of MRFV PRO symmetry mates at two different angles with C-term β-sheet shown in deep purple. Figures were generated in PyMOL (209).

Table 3.3 Crystallographic statistics for MRFV PRO and MRFV PRO-Ub structures.

Table 2. Crystallographic statistics for MRFV PRO and MRFV PRO-Ub structures.

	MRFV PRO	MRFV PRO-Ub
X-ray source	Rigaku R-AXIS IV++	Rigaku R-AXIS IV++
Crystal geometry		
Space group	$P2_1$	I4
Unit cell (Å)	*a*= 43.20 *b*= 73.26 *c*= 54.34; α=β=γ=90°	*a*=*b*= 75.72 *c*= 79.56; α=β=γ=90°
Crystallographic data		
Wavelength (Å)	1.5418	1.5418
Resolution range (Å)	30.90-1.90 (1.97-1.90) *	27.42-2.09 (2.15-2.09) *
Total observations	89460 (5639)	193427 (15741)
Unique reflections	27471 (1718)	13345 (1109)
Multiplicity	3.4 (3.3)	14.5 (14.2)
Completeness (%)	98.9 (99.6)	100 (100)
R_{merge}	0.070 (0.41)	0.091 (0.36)
CC1/2	0.99 (0.74)	0.99 (0.97)
I/σI	7.03 (2.06)	26.3(8.0)
Wilson B-factor (Å^2)	22.91	14.5
Refinement statistics		
Reflections in test set	2366 (235)	1296 (138)
Protein atoms	2175	1683
Ligands	0	16
Solvent molecules	169	239
R_{work}/R_{free}	0.20/0.24	0.20/0.26
RMSDs		
Bond lengths/angles (Å/°)	0.007/0.84	0.002/0.55
Ramachandran plot		
Favored/allowed (%)	97.23/2.42	97.65/1.41
Average B factor (Å^2)	25.77	22.63
Macromolecules	25.34	21.76
Ligands	-	27.96
Solvent	31.38	28.42

*Values in parentheses refer to the highest resolution shell

its neighbour and appears to complete the active site of the neighbouring molecule (Fig. 3.7E). The C-term tails of each symmetry mate jut into the cleft that exists between helices α3 and α7 as well as strand β5 of their neighbour (Fig. 3.7E). The catalytic C61 and incoming His144 residue of the neighbouring symmetry mate arrange themselves to within a few angstroms of each other and appear coordinated for catalysis (inset to Fig. 3.7E (right)). The active site arrangement of the crystallographic symmetry mates more closely resembles the structured active site of TYMV PRO (inset to Fig. 3.7E (left), TYMV in lavender). Finally, surface representations (Fig. 3.7F) reveal the close association of the two MRFV PRO molecules that comprise Figs. 3.7 D & E. The midsection of the C-terminal tails nestle themselves into deep grooves formed between helices α3 and α7 of domain II, as well as strand β5 of the sole β-sheet of domain III. The ends of the C-term tails seamlessly pack up against the convex face of the β-sheet (purple). MRFV PRO is monomeric in solution according to size-exclusion chromatography and this dimeric interaction between the monomers of the ASU is most likely a crystallographic artifact, though the nature of the flexible C-term tail is curious to speculate about.

3.3.6 Comparing the proteases of MRFV and TYMV

MRFV PRO and TYMV PRO share a similar three-dimensional fold (Fig. 3.8A) outside of variability in loops. Loops often they play a large role in protein-protein interactions and substrate recognition (275), so it follows that different substrate/interaction requirements would manifest in loop variability between the MRFV and TYMV proteases. The MRFV PRO structure was compared to known 3D protein structures using the DALI server (276), which revealed TYMV PRO to be the closest structural homologue, with a Z-score of 20.4 (PDB code: 4A5U). The next closest structural homologue was the ovarian tumor domain-containing protein 3 (OTUD3) from

Homo sapiens with a Z-score of 5.9 (PDB code: 4BOU), followed by OTUD1 from *Saccharomyces cerevisiae* with a Z-score of 5.5 (PDB code: 3C0R). TYMV PRO itself has been characterized as a viral OTU DUB based on its overall core fold (25, 172) and appears to have more homology with OTUD3 and OTUD1 with Z-scores of 7.6 and 7.4, respectively. The difference in Z-scores of each PRO domain with ORTUD1/3 can primarily be attributed to large variations in the C-termini of MRFV and TYMV PRO. As mentioned previously, the C-term of MRFV PRO appears to be flexible compared to TYMV. For this reason, TYMV PRO has a more organized active site with its catalytic Cys783 and His849 residues in close coordinating distance, even in the absence of Ub substrate (Fig. 3.8B). Despite their differences, the structural homology of MRFV PRO shared with TYMV PRO, OTUD1 and OTUD3 clearly classifies it as a viral OTU DUB.

Based on sequence alignments and the TYMV PRO structure, the potential flexibility of the MRFV PRO C-terminus may be an unusual feature among marafivirus PRO domains. For TYMV PRO, there is a region at the C-term upstream of the catalytic His known as the "GPP flap" (outlined in black, Fig. 3.2B). This motif has been shown to be essential for the protease to toggle between endopeptidase and DUB activity (277). Mutations in the region decreased DUB activity but did not appear to hinder polyprotein processing at the PRO|HEL junction, indicating the importance of the flap to modulate Ub-dependent antiviral responses but not polyprotein processing (277). Interestingly, aside from the terminal proline (P142) MRFV PRO lacks a complete GPP motif (Fig. 3.2B and Fig. 3.8B). Glycine and proline are key residues in β-turns, and their absence may explain why the C-term of MRFV PRO fails to fold back to form a four-stranded stranded β-sheet following β5 as seen with TYMV PRO (depicted with arrow).

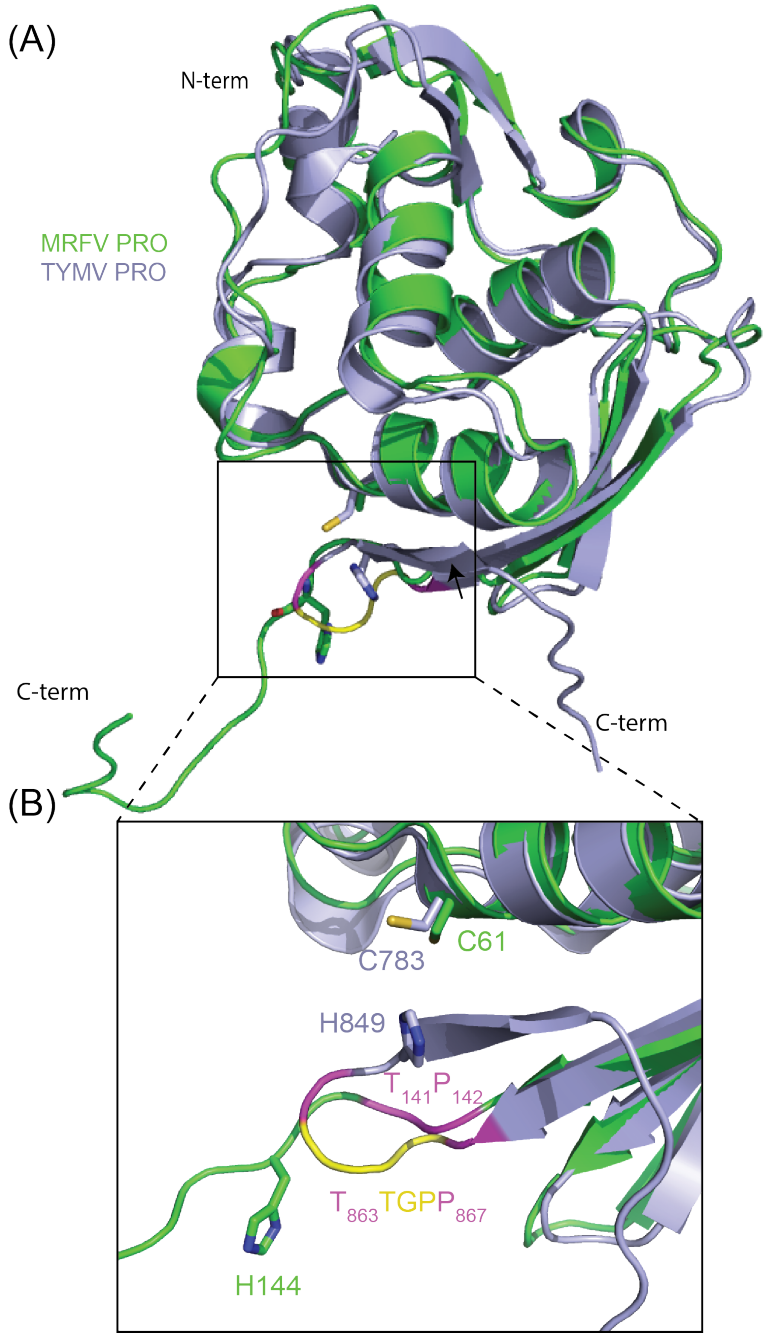


Figure 3.8 Crystal structure of MRFV PRO superposed with TYMV PRO.

Overview *(A)* and close-up *(B)* of MRFV PRO overlayed with TYMV PRO (PDB code: 4A5U). MRFV PRO is shown in green and TYMV PRO is shown in lavender. T141 and P142 (MRFV) and the corresponding T867 and P867 (TYMV) are shown in magenta. The TYMV PRO loop composed of $G^{864}P^{865}P^{866}$ is shown in yellow. Figures were generated in PyMOL (209).

Interestingly, with the exception of MRFV PRO all the marafivirus PRO domains in our study appear to contain a loop region similar to the GPP loop of TYMV PRO, yet we found them to be fully capable of cleaving the viral polyprotein and poly-Ub chains. Indeed, as described below, it could be that the interactions of the enzyme with Ub prompts the C-terminus of MRFV PRO to adopt a conformation that brings H144 close to C61 and generate a complete active site within a single monomer of MRFV PRO that would turnover Ub.

3.3.7 The structure of MRFV PRO in complex with Ubiquitin

MRFV PRO was covalently linked to Ub. Specifically, Ub-3Br is a suicide substrate of deubiquitinating enzymes that is a modified form of WT Ub in that its C-terminus is modified to harbour a reactive C-terminal tail, which can irreversibly bind to the active site cysteines of DUBs through a covalent linkage. The covalently linked protein complex was crystalized in space group I4. Fig. 3.9A shows a remarkably large binding interface between MRFV PRO and Ub. The C-term LRGG tail of Ub nestles deep into the active site channel of PRO, which forms numerous additional interactions with the beta-grasp fold of Ub (Fig. 3.9A & B). As determined through the PISA server (278), there are 38 residues of PRO that are involved with interactions with Ub, which is over 25 % of the residues, covering 948 $Å^2$ (~13 %) of the accessible surface area.

Ub has four key surface features that are typically recognized by DUBs and Ub-binding proteins (168, 279). Two of these features are essential in how PRO recognizes Ub, which are the hydrophobic Ile44 and Ile36 patches (Fig. 3.9A-C). The Ile44 patch is composed of residues L8, I44, H68 and V70. Fig. 3.9B shows the interactions that occur at the Ile 44 patch between PRO and Ub. Four key residues of PRO partake in the stabilizing of Ub on the surface of PRO. Hydrophobic residues P37 and A40 of the long loop connecting α1 and α2 grip against the

hydrophobic Ile44 patch. The pyrrolidine side chain of P37 quite efficiently burrows into the groove in Ub created by I44 and V70. V125 of PRO also uses its hydrophobic isopropyl side chain to facilitate interactions with the lobe of Ub that arises from L8. Although not hydrophobic, R27 is a key residue for stabilizing Ub binding at this region since its guanidino group hydrogen bonds with Ub's main-chain carbonyl group of G47 (Fig. 3.10A). This hydrogen bond orients the guanidino group to press up against the I44 patch and impart another degree of stabilization. Interestingly, only MRFV and OBDV have Arg residues at this position (Fig. 3.2B, triangle at position ~30). R27, along with A45 (Fig. 3.2B, triangle at position ~50), are the only two residues exclusively shared by OBDV and MRFV; however, A45 has no interactions with Ub. R27 is clearly important in MRFV PRO recognition of Ub and could be a factor that we can attribute into MRFV and OBDV lacking the ability to recognize and subsequently process Ub-K^{63} chains.

The Ile36 patch of Ub also has many interactions with PRO as seen in Fig. 3.9C. V125, F118 and S120 are all directly involved with hydrophobic contacts with the Ile36 patch, helping to stabilize Ub on the surface of PRO while also facilitating entry and guidance of the C-term tail of Ub down into the channel that terminates with the active site. V125 is involved with hydrophobic interactions with both patches, showing its importance in recognizing Ub. Similarly, TYMV PRO has an Ile847 (highlighted by a triangle in Fig. 3.2B) at this position and previous studies have shown the importance of this residue in DUB activity (172). Finally, T102 of MRFV PRO forms two hydrogen bonds with L73 of Ub. The first is between the main-chain amide of

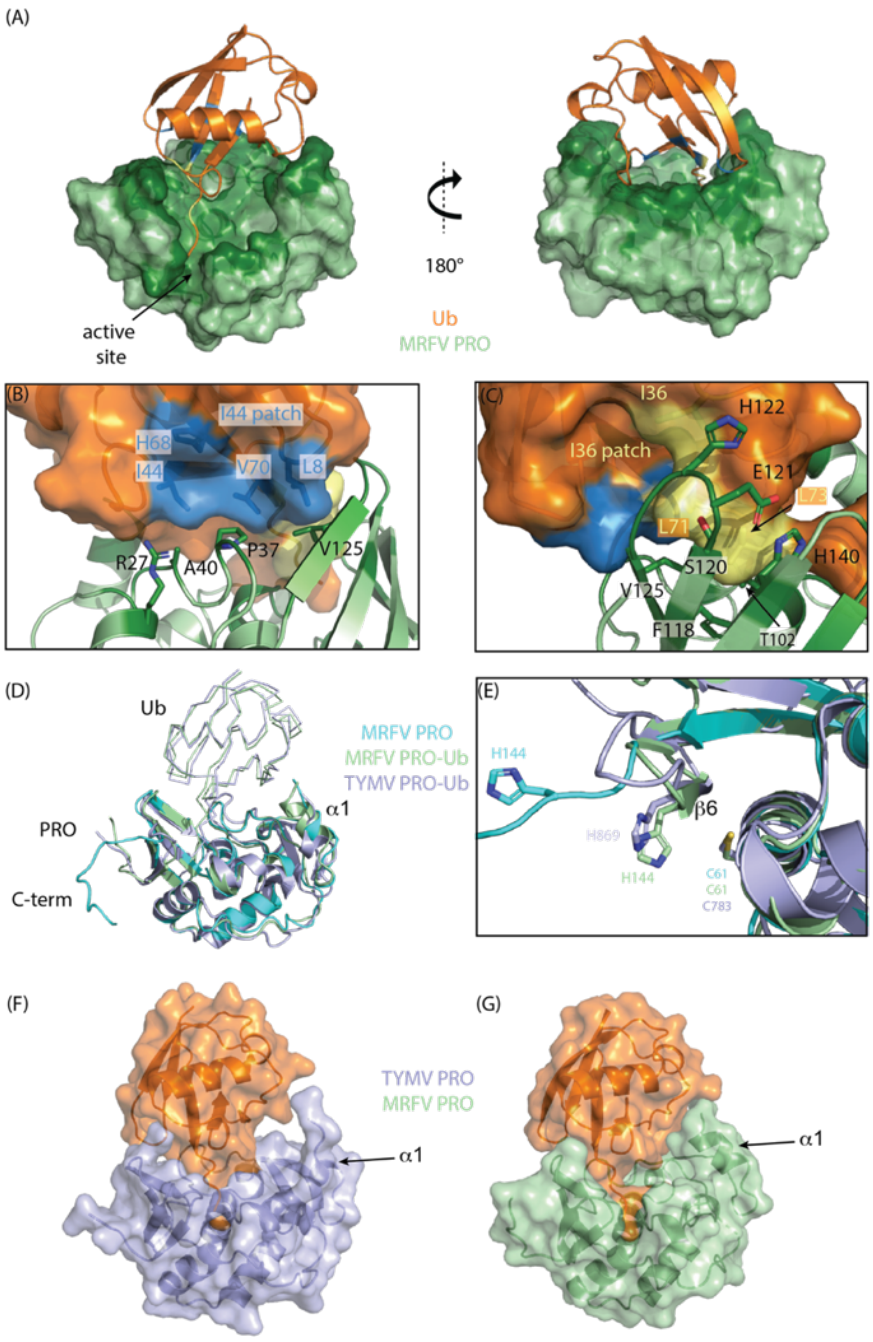


Figure 3.9 Crystal structure of MRFV PRO bound to Ub.

(A) MRFV PRO is shown as a surface representation in green and Ub as a cartoon in orange. Dark green regions are interacting residues of PRO. *(B)* Interactions of the Ile44 patch of Ub (shown in blue) with key residues of PRO. *(C)* Interactions of the Ile36 patch of Ub (shown in yellow) with key residues of PRO. *(D)* Superposition of MRFV PRO, MRFV PRO-Ub and TYMV PRO-Ub (PDB code: 6YPT). Ub molecules are shown in ribbons and PRO domains are shown as cartoons. *(E)* Superposition of the active sites of MRFV PRO, MRFV PRO-Ub and TYMV PRO-Ub. *(F & G)* Surface representations of TYMV PRO-Ub and MRFV PRO-Ub. Figures were generated in PyMOL (209).

T102 and the main-chain carbonyl of L73. The second is between the side chain hydroxyl of T102 and the L73 main-chain amide (Fig. 3.10B & C). H122 of PRO forms a hydrogen bond between $N^{\delta 1}$ of its side chain and the side chain amino group of Q40 of Ub (Fig. 3.10D). Lastly, H140 of PRO forms a critical hydrogen bond between its side chain $N^{\varepsilon 2}$ and the main-chain carbonyl of R74 of Ub (Fig. 3.10E). Although not all of these interactions are directly with Ile36 patch residues, they are key in this general vicinity and help stabilize Ub globally when considered together. Further, the interactions of T102 and H140 interact directly through hydrogen bonding with the "LRGG" tail of Ub, suggesting that they also participate in polyprotein substrate recognition.

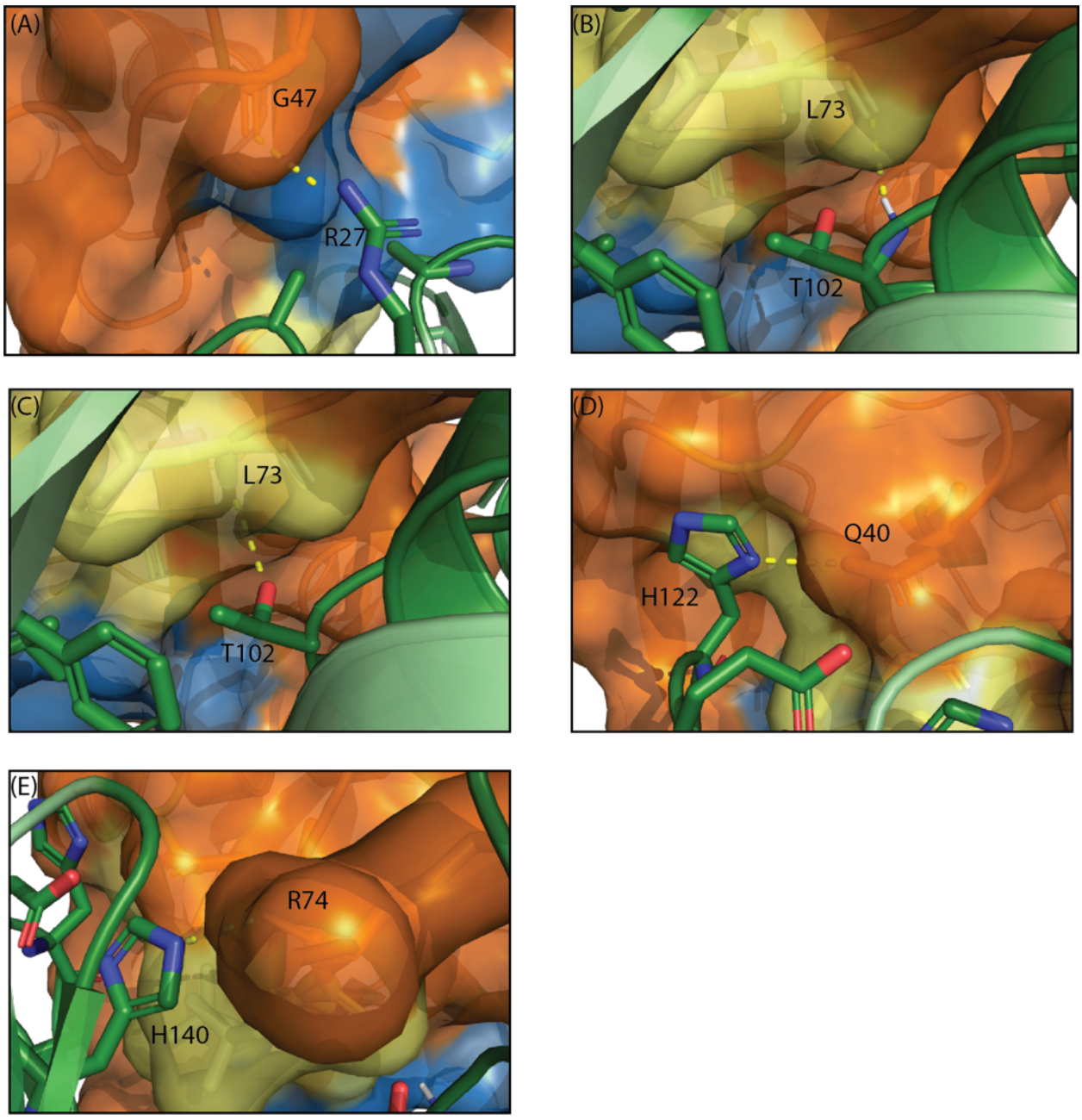


Figure 3.10 Hydrogen bonding interactions of MRFV PRO and Ubiquitin.

MRFV PRO in green and Ubiquitin in orange. Hydrogen bonds are indicated as dashed yellow lines and are those that are directly involved in Ile36 (yellow) and Ile44 (blue) patch recognition and are outlined in the main text. Figures were generated in PyMOL(209).

Overall, the crystal structures of MRFV PRO-Ub and the structure of TYMV PRO-Ub determined in 2020 (PDB code: 6YPT (280)) are similar (Fig. 3.9D). The core secondary structure and folds are maintained with only subtle variation within loop regions. Ub binds to both proteins in a similar orientation. Interestingly, in the presence of Ub, the C-term tail of MRFV PRO adopts a conformation highly similar to the GPP flap of TYMV PRO (unliganded and liganded) as there now is a turn following β5, which allows for the formation of a 4-stranded β-sheet and a more canonical, rigid active site (Fig. 3.9D & E). Further, MRFV PRO forms a much more extensive interaction with Ub compared to TYMV PRO, with a complex formation significance score (CSS) of 1.000 (scale being 0 to 1) as determined from the PISA server (278) (Fig. 3.9F & G). TYMV PRO-Ub has a CSS of 0.822. Much of this is due to α1 of MRFV PRO, which is more fully formed in the MRFV PRO-Ub structure compared to TYMV PRO-Ub and may also contribute to poly-Ub chain type specificity. Interestingly, MRFV PRO (along with SARS-CoV-2 PLpro) have the highest CSS score of viral DUBs that have been structurally characterized bound to Ub, which includes Crimean-Congo Hemorrhagic Fever Virus, Equine Arteritis Virus, Murine Cytomegalovirus, Dugbe Virus, Hazara Orthonairovirus, Mouse Hepatitis Virus, Middle East Respiratory and Severe Acute Respiratory Syndrome Coronaviruses.

3.3.8 Preliminary data of OBDV PRO bound to Ubiquitin

Purified OBDV PRO was covalently coupled to Ub identically to MRFV PRO to form a OBDV PRO-Ub complex (Fig. 3.11A & B). The complex was highly stable and could readily be concentrated to concentrations > 45 mg/ml. Due to its high stability and solubility, the complex did not crystalize or even precipitate in hundreds of conditions that were screened at concentrations far higher than what was done for MRFV PRO-Ub. Nearly a year after their initial setup, crystals

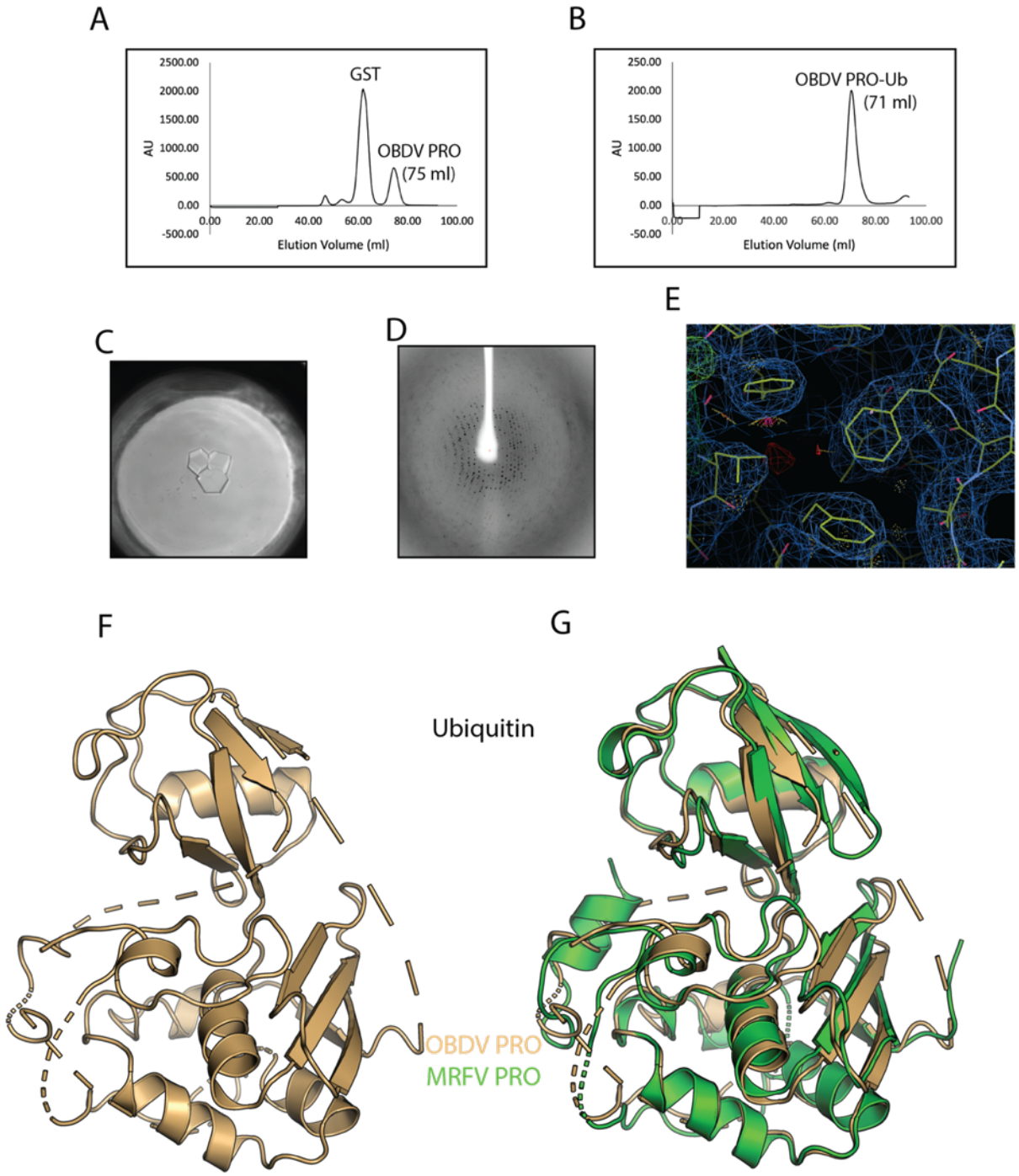


Figure 3.11 Purification and preliminary structure determination of OBDV PRO-Ub.

A & B) Purification of OBDV PRO and OBDV PRO-Ub by gel filtration. C & D) Crystals of OBDV PRO-Ub and diffraction data. E) Initial electron density map from PHENIX AutoBuild. F & G) Preliminary structure of OBDV PRO-Ub (gold) and superposition with MRFV PRO-Ub (green), chains breaks showed with dashed lines.

of the complex were found in drops that were setup at 30 and 35 mg/ml (Fig. 3.11C) and diffracted out to a resolution of ~3.2 Å, so data was collected in-house (Fig. 3.11D). Phasing was done using a multicomponent molecular replacement search using the structure of MRFV PRO from the MRFV PRO-Ub structure, as this conformation of MRFV PRO likely more closely resembles that of what OBDV PRO would adopt, as they are both substrate bound. The structure of Ub (PDB 1UBQ) was used as the other component for the MR search. PHENIX phaser (253) was successfully able to phase the data and the final translation function Z-score was 27.1, significantly higher than 8, which is thought of as being the cut-off value for whether PHASER has successfully solved a structure. Model building was carried out using PHENIX AutoBuild (253), which generated electron density maps that fit well with the structure built, although they are relatively weak density (Fig. 3.11E). A preliminary structure was determined that had significant chain breaks in both PRO and Ub that will have to be worked on (Fig. 3.11F). Regardless, the structure superposes very well with that of MRFV PRO-Ub, although the active site was not able to be built using AutoBuild, which suggests potential differences between the two structures. As can be seen in Fig. 3.2B, OBDV PRO has many more residues N-term to the catalytic His in comparison to MRFV PRO, which is potentially why the active site region was difficult to construct by AutoBuild; however, this preliminary data is promising and should shed more light onto the story of how marafiviral proteases recognize Ub and how that compares between members in the genus as well as family.

3.3.9 Use of ubiquitin variants as competitive inhibitors

The strategy of using UbVs as competitive inhibitors has shown effective against a few RNA viral proteases (246). The strategy relies on exploiting the generally weak affinity of the

active sites of these enzymes to Ub and incorporating amino acid substitutions and/or deletions throughout WT Ub to create variants of Ub that have an overall higher avidity towards the enzyme to act as competitive inhibitors, be it via DUB activity and/or polyprotein processing. This portion of this project is still in its relative infancy, although a fair number of achievements have been made.

As above mentioned, UbVs towards CSDaV PRO were initially generated using a phage display library, where the most potent UbV bolstered an EC_{50} of ~8 nM towards the enzyme, which is several orders of magnitude higher than seen with WT Ub. This UbV did bind efficiently to PRO in attempts to purify the complex for crystallization, however, there was always a state of equilibrium at play in which not all of the enzyme would be coupled with the UbV even in > 5-fold molar excess concentration of the UbV, suggesting that this UbV was not as effective at binding the enzyme as previously thought.

Fortunately, we were able to generate UbVs towards MRFV PRO with EC_{50} values of ~5 nM, the most potent of which was termed MRFV UbV2. This was a more promising target to work with as there is greater agricultural relevance in North America and a more appealing aim, as this is the type member of the genus. Luckily, this UbV bound to PRO very efficiently and completely, even in only slight molar excess concentrations (Fig. 12A & B). The protein complex formed crystals in several conditions with the most promising hits coming in a condition composed of 0.2 M sodium acetate trihydrate and 20 % PEG 3350 and 15 mg/ml protein. All conditions that yielded crystals led to crystals that were often very thin plates that were actually many crystals stacked on top one of another (Fig. 3.12 C top), which did not lead to very great diffraction and made it difficult to isolate individual crystals. Using the initial crystallization condition

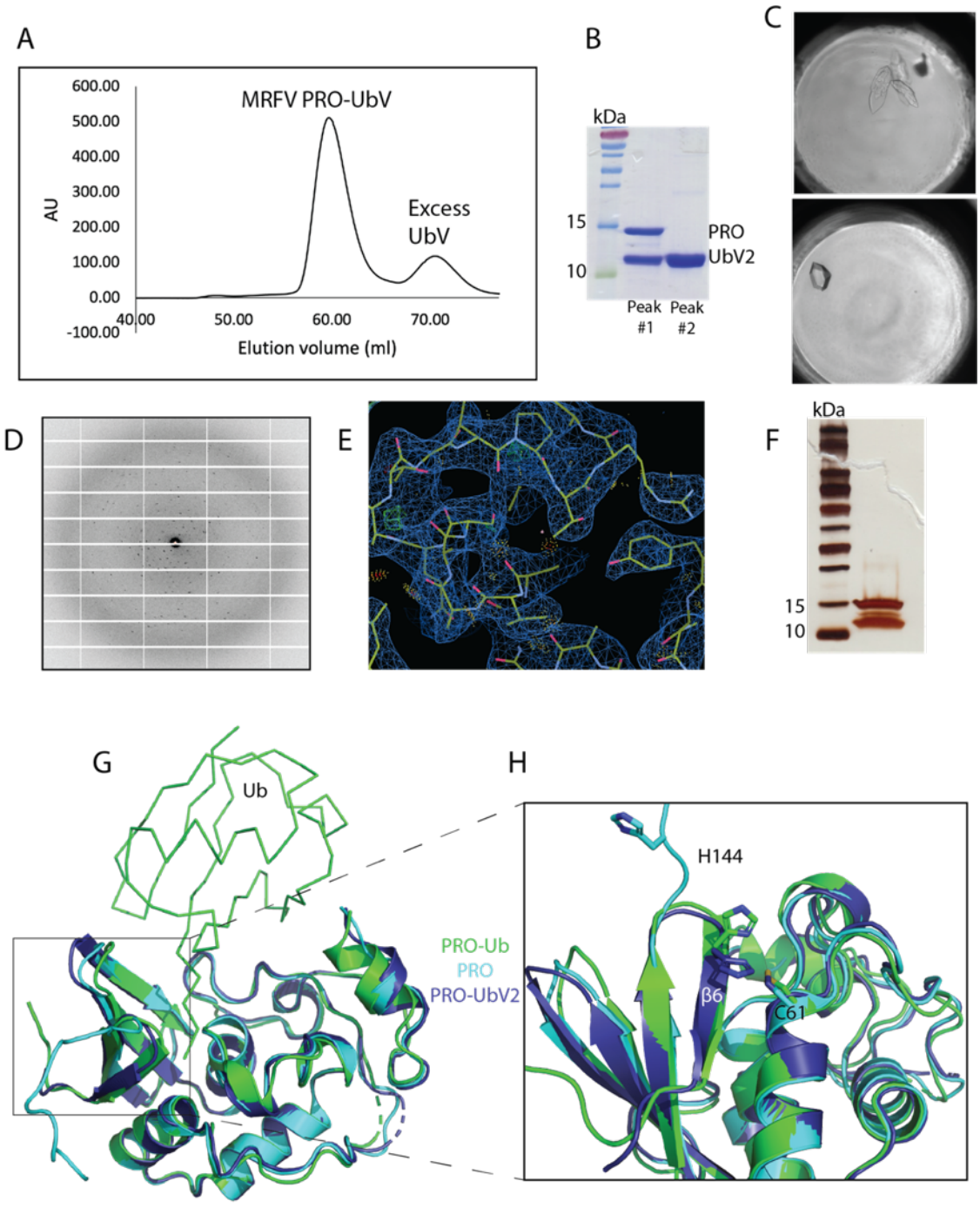


Figure 3.12 Crystallization of MRFV PRO bound to UbV2.

A & B) Purification of the MRFV PRO-UbV2 complex by gel filtration. C) Crystals of the protein complex, conditions described in text. D) Diffraction data of crystal from (C); bottom. E) Electron density map from the data collected in (D). F) Silver-stained SDS-PAGE gel of crystals. G) Superposition of MRFV PRO (cyan), MRFV PRO-Ub (green), and MRFV PRO-UbV2 (dark blue). H) Closeup of superposition with active site residues labeled.

supplemented with 10 mM spermine tetrahydrochloride yielded the most promising crystal (Fig. 3.12C, bottom), which we were able to collect data from out to ~2.2 Å (Fig 3.12D). Unfortunately, problems arose during phasing in which we were never able to come up with a solution that contained an appropriately placed UbV molecule(s) in acceptable density, even though a multicomponent MR search using MRFV PRO (no substrate structure) and Ub (PDB: 1UBQ) scored a translation function Z-score of 15, indicating that PHENIX Phaser had solved the structure. Regardless, the electron density maps for PRO were quite good (Fig. 3.12E). The structure of PRO from UbV2-bound crystals resembled the previous structures determined herein (Fig. 3.12G, dark blue). What was interesting to see was the C-term tail more closely resembles that of the Ub-bound structure (Fig. 3.12H) and forms the same β6 strand, which leads us to believe that it may also be in a substrate bound conformation with the UbV present. Further, the active site is even more closely formed than in the Ub-bound structure (shown in green) with His144 and Cys61 being the closest in distance of all three structures.

The crystals were never able to be reproduced in the same form as the one that the data was collected from. They readily formed but remained in their stacked thin plate arrangement, making it difficult to get meaningful data from. Selenomethionine derivatized protein was produced and put through crystal screens (both new and the original condition), however crystals of high enough quality for data collection were never obtained. To confirm if the UbV was actually present in crystals that form in the original condition (albeit in the plate form), crystals were harvested and washed thoroughly and solubilized and run out on an SDS-PAGE gel, which was subsequently silver-stained (Fig. 3.12F). This revealed the presence of both proteins within the crystals. This venture is absolutely worth pursing in order to determine at the molecular level why the UbV binds with much higher affinity to PRO in comparison to WT Ub. There is promise to push forward, as

the complex seems to crystalize quite easily, though will require more attention and with special focus on selenomethionine derivatized protein.

3.3.10 Development of transgenic Z. mays

The MRFV UbV2 coding sequence was inserted into a binary vector (pDW3596) for *Agrobacterium*-mediated transformation into maize line B104. This variety of corn was chosen over the other line offered by the plant transformation facility at Iowa State University (Hi II), as the seeds are bigger. The larger seed harbours a larger scutellum and embryo regions on the outer seed, which are the targets for vascular puncture inoculation using a linear-motion tattoo machine to actually infect the seed and thereby the subsequent plant with virus, which is the most efficient way to infect *Z. mays* in a lab setting with MRFV (139). The binary vector contains the coding sequence for MRFV UbV2 with an N-terminal 3xFLAG tag, which have been codon optimized for *Z. mays* and whose constitutive expression will be driven by the maize ubiquitin 1 (Ubi-1) promoter. Once constructed and sequenced, the vector was sent off to Iowa State University where they were able to use their *Agrobacterium* system. The transgenic seed was developed after nearly 300 days and has been sent to our collaborators at the USDA (Fargo, ND). Currently they are working to make the corn plants homozygous for the UbV transgene via self-pollination. We have recently received tissue samples of progeny plants, which will be used to assay for the transgene by western blot (anti-FLAG) as well as PCR of maize genomic DNA with primers specific to the transgene. Following confirmation of progeny harbouring the transgene, more self-pollination will be carried out to establish a fully homozygous plant.

3.4 Conclusion

Together, our results provide new structural and biochemical insights into the papain-like cysteine proteases present in the polyproteins of marafiviruses. We demonstrate for the first time that these enzymes have deubiquitinating activity in addition to acting as endopeptidases that process the viral polyprotein. *In vitro* and *in cellulo* polyprotein processing assays have provided first insights into how marafiviruses process their polyproteins, which appear to have features distinct from their tymovirus relatives. Our structural findings of the maize-affecting type member MRFV reveal that the enzyme has regions that appear to be quite dynamic, which assist in recognizing different viral and cellular substrates. Unexplored nuances exist that can still be investigated to understand how certain marafaviral DUBs selectively process only K^{48} poly-Ub chains and not K^{63} poly-Ub chains. Structural analysis also reveals that MRFV PRO has one of the most extensive interaction surfaces with Ub. Collectively, these results lay the groundwork in biochemically understanding this class of DUBs and sets the stage for future studies to exploit these enzymes, as is our direction with Ubiquitin variants, which our lab is now actively working with in collaboration with plant biologists to ideally make transgenic plants that are less susceptible to viral infection.

3.5 *Future directions*

It would be interesting to introduce mutations into MRFV and OBDV PRO at R27 or R34 (MRFV and OBDV, respectively) since these two marafiviral PRO domains are the only two with an Arg at this position. As seen in the crystal structure of MRFV PRO-Ub, this residue is important in recognizing Ub. It could also be grounds for differential recognition of poly-Ub chains and could explain MRFV and OBDV PRO recognition of only K48 chains and not K63 chains. Another interesting mutation to include would be an insertional mutation in MRFV PRO at the corresponding region of TYMV PRO's "GlyProPro" flap, which is mostly absent in MRFV PRO. Including the Gly-Pro insertional mutant and assessing activity of the enzyme towards Ub-AMC and other substrates could shed light on differences between these proteases in how they recognize ubiquitin substrates. It would also be beneficial in construct a peptide representing the Hel|RdRp junction (LRGA) of the MRFV polyprotein. This could be done with an LRGA-AMC peptide or an internally quenched fluorescent peptide and performing Förester resonance energy transfer assays to confirm if this is also a cleavage site of the polyprotein.

It may also be useful to re-establish the BY-2 cell line and launch an efficient protoplast assay system. These cells are much easier to work with as they require very little maintenance and will divide indefinitely if managed appropriately. The amount of labour would be far less than with the *A. thaliana* system. It would be interesting to carry out localization assays of the various components of the MRFV major polyprotein. Antibodies could easily be generated against PRO, now knowing the structure of the enzyme to determine a surface exposed epitope of the enzyme. Peptides could be designed towards the Hel and RdRp domains based on previous methods for TYMV (163, 165). It would be curious to see if the domains typically associate with chloroplast

membranes as is the case for TYMV and if the PRO domain also drives localization to these regions of the cell.

4 Chapter 4: Probing the deubiquitinating activity of benyvirus and SARS-CoV-2 papain-like proteases

Contributions statement:

Benyvirus data:

All benyvirus work was performed by Ankoor Patel following discussions with Dr. John Weiland (USDA, Fargo)

SARS-CoV-2 data:

This work was largely conceptualized by Drs. Sachdev Sidhu (U of Toronto), Roman Melnyk (U of Toronto) and Brian L. Mark (U of Manitoba). UbV selections were carried out by members of the Sidhu lab. Recombinant protein purification and crystallization presented in this chapter were done by Ankoor Patel and Cole Slater (Technician, Brian Mark lab).

4.1 Introduction

4.1.1 Benyviruses overview and PLCP

Members of the family *Benyviridae* are another group of plant-affecting viruses with +ssRNA genomes with four members in the genus *Benyvirus* (281). The type member of the genus is the beet necrotic yellow vein virus (BNYVV), which has the closest sequence and genomic architecture similarity to its genre member, the beet soil-borne mosaic virus (BSBMV) (281). BNYVV has been studied for decades and is found throughout the globe, being identified in every region of the United States where sugar beets (*Beta vulgaris*) are produced as a source of refined sugar (282). Symptomatically, the virus can be identified in infected plants that exhibit mosaic patterns in the leaf, along with chlorosis in the veins of leaves (282). The most problematic symptom of infection is the onset of rhizomania, which is when the roots grow secondary root structures that are dead and unable to uptake water, resulting in severe taproot stunting and ultimately decreased plant mass, which is problematic for sugar beet farmers as this manifests in reduced sugar yield (282). The vector for these viruses are slime molds, specifically *Polymyxa betae*, which is ubiquitously found in soils and also are pathogens for *B. vulgaris* (282).

The genome for BNYVV and BSBMV are multipartite and composed of four to five molecules that are 5'-capped with a 3' polyA tail (RNA1-RNA5), which act as mRNA as described for the viruses in previous sections (281, 283). RNA1 and RNA2 are arguably the most important, as they contain the viral replicase and structural proteins, whereas RNA3-5 seem to be important in vector transmission (283). Field viral isolates typically harbour all 4-5 RNA molecules, however, greenhouse isolates that only spread through manual mechanical inoculation typically lose RNA3-5 over time, illustrating that only RNA1 and 2 are necessary for replication in infected plants (282). RNA1, the focus of this section, is ~6.7 kb in length, encodes for one major

polyprotein, and has very similar genomic architecture to the marafiviruses (Fig. 1.3, bottom). The major difference is that the purported cysteine protease and helicase domains have swapped locations in the genome, and thusly the protease sits directly upstream of the RdRp-coding region (281). A study in 1985 using RNA1 in a cell-free system of labelled translation products revealed a protein sharing the same theorized molecular weight of the entire polyprotein as well as smaller proteins that the authors believed were products of proteolytic processing of the major polyprotein (284). Nearly a decade later, a paper showed that in quinoa protoplasts infected with an isolate of BNYVV also generated two potential cleavage products of the major polyprotein, presumably a fusion of MET|HEL|PRO (150 kDa) and the RdRp (66 kDa) alone. These findings are often referenced to in literature; however, they are not fully conclusive and slightly speculative. To date there has been no significant biochemical or structural characterization of these purported PLCPs that also may be OTU-like (13). In this chapter we show that these proteases likely have proteolytic processing abilities specifically as Ub hydrolases, however, if and where these enzymes cleave their polyprotein is still unclear.

4.1.2 SARS-CoV-2 and PL^{pro}

As it has become increasingly apparent, the severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) and other related coronaviruses have the potential for incredibly detrimental epidemiological outcomes, having been responsible for a global pandemic unlike anything that has been seen in over a century. SARS-CoV-2 is also in the order *Nidovirales* along with PRRSV, although it has a much larger genome at ~30 kb in length (285). Like the arteriviruses, coronaviruses also produce two major polyproteins pp1a and pp1ab identically to how described with PRRSV, with a PRF event at the junction of the coding regions of the two major ORFs (285).

Unlike PRRSV, SARS-CoV-2 only has two proteases involved with cleaving pp1a and pp1ab, the $3CL^{pro}$ (chymotrypsin-like cysteine main protease) and PL^{pro} (PLCP) (285). PL^{pro} is encoded in the nsp3 coding region of ORF1a and cleaves at the nsp1|2, nsp2|3, and nsp3|4 junctions, where $3CL^{pro}$ has 11 total cleavage targets (286).

As is the case with non-main proteases of +ssRNA viruses, PL^{pro} from related coronaviruses and SARS-CoV-2 have auxiliary function in host innate immune system modulation by acting as deubiquitinating and de-ISGylating enzymes (188, 268, 287, 288). ISG15 is one of the most integral signalling protein in response to pathogen invasion and, although SARS-CoV and CoV-2 PL^{pro} are ~83 % identical at the amino acid level, their host substrate preference is surprisingly different in respect to Ub and ISG15 (268). It was shown that SCoV2 PL^{pro} has a much higher affinity for ISG15 cleavage than Lys48-linked di-Ub, the opposite of which was shown for SCoV PL^{pro} (268). Both enzymes have nearly identical structures, although key differences in their "S2 sites" are present (268). The S2 site is a region of the enzyme that recognizes the second molecule of Ub in a di-Ub chain attached to a substrate, specifically the molecule that is not covalently attached to the target (268, 289). The structure of one molecule of ISG15 is highly similar to a molecule of di-Ub with a K48 linkage, where ISG15 can easily be seen as two separate domains similar to K48 di-Ub (290). Like Ub, The C-term domain of ISG15 also ends in an "LRGG" motif that is covalently tethered to a cellular substrate protein's Lys residue (291). It is the N-term domain of ISG15 (N-term Ub-like fold) that mimics the non-covalently-linked molecule of Ub in a di-Ub chain, both of which are recognized by the S2 site of PL^{pro}, however, certain residue differences in the S2 site give SCoV2 PL^{pro} a higher affinity for ISG15 over di-Ub in comparison to SCoV PL^{pro} (268, 289).

The search for treatments against COVID-19 has been ground-breaking in its speed and proficiency, with vaccines being the primary and most efficacious mode of prevention (292). It does follow, though, that alternative means of treatment should and have been explored, such as drugs that target the RdRp, $3CL^{pro}$ and PL^{pro}. Several antivirals have been implicated in effectively inhibiting PL^{pro} and should still be explored for potential future coronavirus outbreaks (268, 293–296). The work presented in this chapter involves using the UbV strategy with an alternative purpose as a means to screen for small molecule inhibitors of PL^{pro} which will be discussed in the sections to come.

4.2 Materials and Methods

4.2.1 DNA Constructs

Table 4.1 Primers used for DNA constructs

Oligo name	Sequence (5' to 3')	Purpose
BNYVV_PRO_F	GATATA**GGATCC**GTTGTTGTTCTGG	BNYVV PRO in pGEX
BNYVV_PRO_R	TATATC**CTCGAG**TTACGCCGCG	"
BSBMV_PRO_F	GATATA**CATATG**CATCACCATCACCATCACTCG	BSBMV PRO in pET24b+
BSBMV_PRO_R	TATATC**CTCGAG**TTAGCCGCCTGCTAAGTTC	"
PLPRO_F_pGEX	GATATA**GGATCC**GAAGTGCGTACTATC	PLpro in pGEX
PLPRO_R_pGEX	TATATC**CTCGAG**TTACTTAATGGTGGTTGTATAG	"
PLPRO_F_pET24	GTATAT**CATATG**GAAGTGCGTACTATCAAAGTATTTACG	PLpro in pET24b+
PLPRO_R_pET24	TATATC PLPRO_F_pET24TTACTTAATGGTGGTTGTATAGGAATTTTC	"

*Restricting sites in bold

Synthetic DNA coding for BNYVV PRO was designed based on the purported PRO domain coding sequence on UniProt (Q65667, residues 1285-1388 of the main polyprotein from RNA1). The domain was also designed based on a sequence alignment with the construct of MRFV PRO that was crystalized herein and was selected to be as compact as possible while retaining the active site residues (Fig. 4.1A, top). Synthetic DNA was codon optimized for expression in *E. coli*, PCR amplified and ligated into pGEX-6p-1 using standard restriction cloning. Synthetic DNA coding for BSBMV PRO was designed similarly, although was fully C-terminally extended to

include its cleavage site at the RdRp junction (Fig. 4.1A, bottom). Construction was done identically to BNYVV PRO, but the insert was placed into the pET24b+ expression vector with an in-frame C-term 6xHis tag. The entire coding sequence for SARS-CoV-2 PL^{pro} was codon optimized for expression in *E. coli* and prepared synthetically, PCR amplified and cloned into pGEX-6p-1 or pET24b+ using standard restriction cloning. The latter was constructed to include an N-term 6xHis tag with a PreScission protease site to follow. All constructs were transformed into *E. coli* BL21(DE3) Gold following Sanger sequence verification for protein expression and purification.

4.2.2 Protein expression and purification

All strains harbouring their respective expression vector were grown overnight in LB containing 150 µg/ml ampicillin (pGEX/pET53-DEST) or 35 µg/ml kanamycin (pET24). The following morning, the overnight cultures were used to subculture fresh LB containing the appropriate antibiotic and grown to an OD_{600} of ~0.8 at 37 °C with shaking, followed by protein induction with 0.5 mM IPTG at 16 °C for ~18 h. PL^{pro} construct-containing cells were also supplemented with 10 mM sterile ZnCl during induction. Cells were harvested the following day and frozen at -80 °C or used immediately for protein extraction.

Protein purification for the GST fusion proteins was done identically as outlined in section 3.2.3 for the marafiviral PRO domains. The only exception was that PL^{pro} was purified in pH 7.5 TRIS-HCl and the cleaved GST domain was recaptured using a fresh batch of GST-bind resin the following day on the dialyzed sample, as the GST and PL^{pro} are too similar in size to be efficiently separated by gel filtration. Concentration was quantified by A_{280} (ε/1000 = BNYVV PRO 18.45 $M^{-1}cm^{-1}$; PL^{pro} 45.27 $M^{-1}cm^{-1}$; BSBMV PRO 39.42 $M^{-1}cm^{-1}$).

Purification of BSBMV PRO in pET24b+ or PL^{pro} UbVs in pET53-DEST was done as mentioned in section 3.2.3 for the His-tagged MRFV PRO-HEL fusion proteins, as was the case for PL^{pro} in pET24b+ with a few changes. The UbVs and PL^{pro} purification buffer was pH 7.5 TRIS-HCl and the N-terminal 6xHis tag of PL^{pro} was removed using in-house prepared PreScission protease during protein dialysis overnight. PL^{pro} was purified away from the affinity tag by gel filtration. PL^{pro}-UbV coupling was done as outlined in 3.2.6.

4.2.3 Deubiquitination assays

Ub-AMC assays were done as detailed in section 3.2.4 using a SpectraMax iD5 microplate reader (Molecular Dimensions). Poly-Ub chain hydrolysis assays were performed as detailed in section 3.2.5. Enzyme and substrate concentrations are detailed in the respective results and discission sections.

4.2.4 Protein crystallization

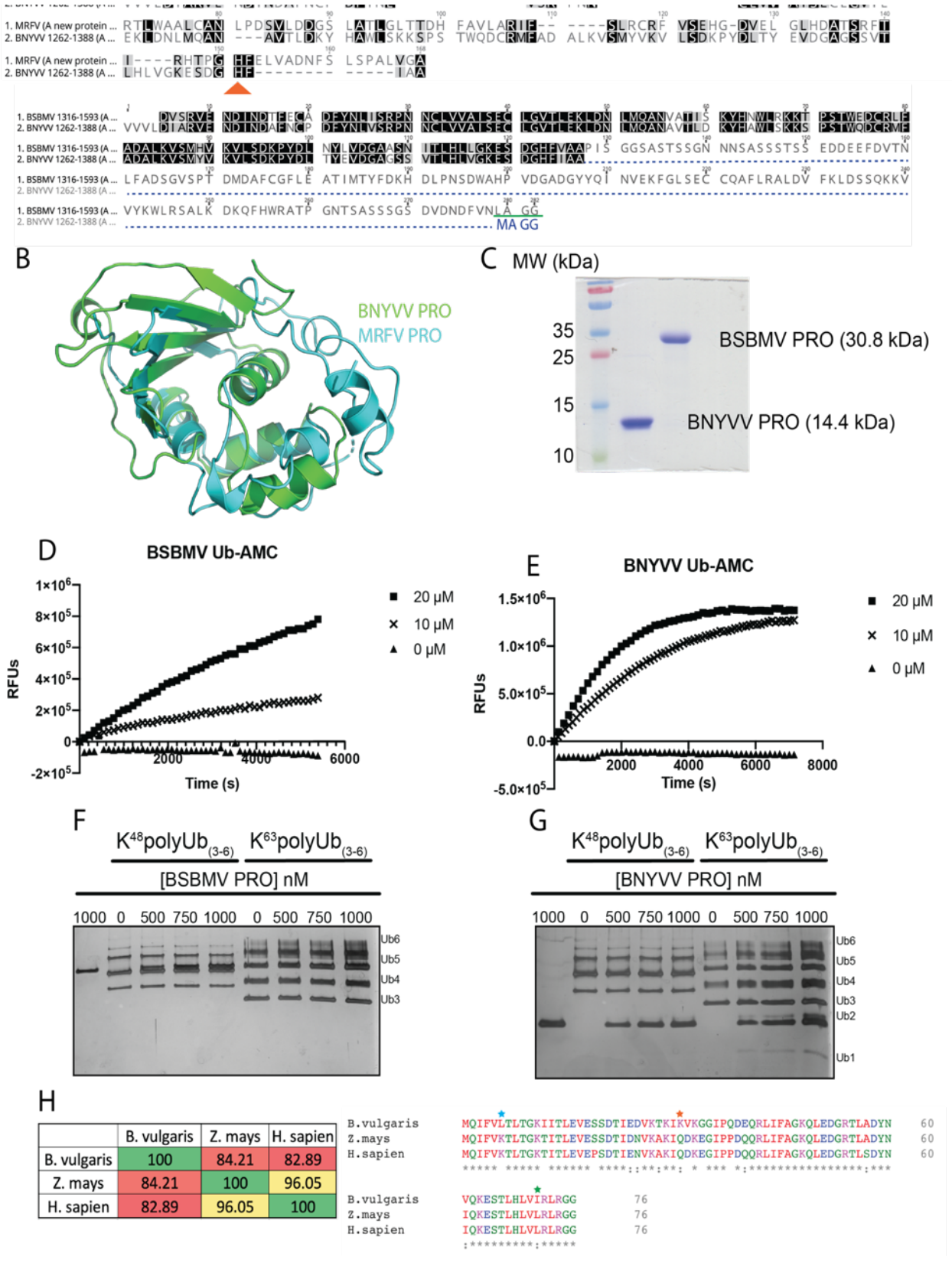
PL^{pro} bound to UbV1was first crystalized at 10 and 15 mg/ml at 4 °C in a sitting drop format using the vapour diffusion method. Crystals appeared after two days in a condition containing 100 mM Bis-TRIS propane (pH 6.5), 200 mM sodium bromide, and 20 % PEG 3350.

4.3 Results and Discission

4.3.1 Construction of the BNYVV and BSBMV protease domains

The BNYVV PRO domain was designed based on a UniProt (Q65667) annotation of the domain within ORF1 of a Japanese isolate of BNYVV. The purported PRO domain spans residues 1285-1388 of the major polyprotein. A sequence alignment with MRFV PRO (Fig. 4.1A, top alignment) revealed relatively low sequence similarity (~25 %) but did expose that the catalytic Cys and His residues are shared (orange triangles). In attempts to make sure the entirety of the protease domain was captured, 23 additional N-term residues were included starting from polyprotein residue 1262 instead of 1285. Using the Phyre2 server (297), a homology model was generated (Fig. 4.1B, green). The model most closely resembled the Hazara virus OTU domain, which was crystalized covalently bound to Ub (PDB 7JMS), followed by the OTU domains of the Erve and Farallon viruses (PDB 5JZE & 6DX5, respectively). Based on the homology model alone, we believed that benyviral PRO domains may also be DUBs. Superposition of MRFV PRO with the BNYVV PRO homology model also showed a promising degree of conservation (Fig. 4.1B).

Interestingly, there is no Ub C-terminal LRGG-like motif to follow directly after the BNYVV PRO domain as is seen with the marafiviruses and their LXG(G/A) motif at the C-term of PRO. There is a downstream MAGG motif ~150 residues that our USDA collaborators believe may be the cleavage site (Fig. 4.1A, bottom alignment); however, these residues were omitted in order to maintain as compact a protein as possible for successful recombinant expression and purification (Fig. 4.1C).



	B. vulgaris	Z. mays	H. sapien
B. vulgaris	100	84.21	82.89
Z. mays	84.21	100	96.05
H. sapien	82.89	96.05	100

Figure 4.1 Exploring the theorized protease domains of two beet-affecting benyviruses.

A) Sequence alignment of MRFV PRO and BNYVV PRO (top). Catalytic Cys and His residues are indicated with orange triangles. Extended BSBMV PRO aligned with truncated BNYVV PRO (bottom). Putative C-term LAGG and MAGG cleavage sites are indicated or annotated. B) Structure superposition of MRFV PRO (cyan) and homology model of BNYVV PRO generated using Phyre2 (297). C) SDS-PAGE gel of Purified BNYVV and BSBMV PRO. D & E) Ub-AMC assays of BSBMV and BNYVV PRO. F & G) Silver-stained 10% TRIS Tricine gels representing Poly-Ub chain hydrolysis assays of BSBMV and BNYVV PRO using Lys48- and Lys63-linked poly-Ub substrates. H) Left; percent sequence identity matrix of Ub from *Beta vulgaris*, *Zea mays* and *Homo sapien*. Right; multiple sequence alignment of Ub amino acid sequences from all three aforementioned species.

4.3.2 Exploring DUB activity of BSBMV and BNYVV PRO

Preliminary Ub-AMC assays with the purified BNYVV protease at low micromolar concentrations (< 5 μM) did not convincingly show Ub hydrolase activity (not shown). Further, a fluorogenic peptide (MAGG-AMC) representing the theorized cleavage target of the polyprotein at the PRO|RdRp junction also did not reveal proteolytic activity.

After discussions with our collaborators, we decided to construct an expression vector encoding for the highly similar BSBMV PRO domain that essentially started at the same N-term position as BNYVV PRO but was extended to include its putative "LAGG" cleavage site (Fig. 4.1A, bottom alignment, underlined in green). This resulted in soluble protein (Fig. 4.1C) that actually showed activity against Ub-AMC (Fig. 4.1D), but in order to see substrate turnover, enzyme concentration had to be nearly double that of what was used for the marafivirus PRO domains. Using these same, elevated concentrations, BNYVV PRO was attempted again, and Ub-AMC hydrolysis was seen (Fig. 4.1E).

Poly-Ub chain substrates were used that were either Lys-48 or -63 linked to test for substrate specificity. BSBMV PRO did not seem to hydrolyze either K48 or K63 chains (Fig. 4.1F) even when enzyme was used at 1000 nM concentrations, which was fivefold excess of the upper limit of concentrations used for the marafivirus proteases. BNYVV PRO also did not hydrolyze K48 poly-Ub chains but did show the beginnings of hydrolysis towards K63 chains, where di- and mono-Ub can faintly be seen in the presence of enzyme (Fig 4.1G, lanes 7-9). K63 poly-ubiquitination in plants is often a signal for vesicular trafficking, plant developmental processes and membrane transport specific to nutrient uptake and distribution (259, 298), though plant-affecting viral DUBs acting towards K63 linkages has not thoroughly been explored as it has in mammalian systems (16, 168, 264).

4.3.3 Differences in Ubiquitin amongst various species

When trying to understand why these enzymes had such poor activity against the various Ub substrates, a closer look into the sequences of Ub between many species was surprising to us. In all three species analyzed (*Zea mays*, *Beta vulgaris* and *Homo sapien*) Ub is a 76 amino acid long protein (Fig. 4.1H, right). Interestingly, Ub from *B. vulgaris* the lowest sequence similarity to human Ub at ~83 % (Fig. 4.1H, left) and ~84 % sequence similarity to maize Ub. Human and maize Ub are quite similar at ~96 %. The Ub substrates that are commercially available are all designed on the sequence of human Ub. As *B. vulgaris* Ub is the least similar in sequence, this could be one reason as to why BNYVV and BSBMV PRO have reduced hydrolysis profiles towards Ub-AMC and the poly-Ub substrates.

A few things of interest emerged when looking at the sequence of beet Ub, such as it lacking Lys6 (Fig. 4.1H, blue star). Lys6-linked poly Ub chains are considered atypical as they occur in quite low abundance in plants, but have been thought of as having some relation to DNA damage signaling (171, 262, 299). It is curious that beet Ub lacks this Lys residue, as these are arguably some of the most important residues within Ub and the fact that it is also conserved in maize. Another interesting finding was that beet Ub seems to have a Lys residue at position 31 (orange star) in place of a Gln residue. It could be possible that this may be another residue specific to beet Ub that is used as a branching point for a different type of poly-Ub chain with tentatively different cellular implications. Although not as stark a difference, beet Ub has an Ile at position 71 in place of maize and human Ub Leu (green star). The only reason this is of note is that Leu71 of human and maize Ub is an integral residue that is a member of the Ile36 hydrophobic patch, which as mentioned above, is a critical region of Ub that is often recognized by viral DUBs, as is the case for MRFV PRO and its recognition of Ub (Fig. 3.9C) (16, 168). Collectively, these differences

and potentially other amino acid differences that have not been pointed out could be attributed to why BSBMV and BNYVV PRO recognize these substrates rather poorly.

4.3.4 SARS-CoV-2 PL^{pro} purification and DUB activity

The goal of this project was to develop small molecule inhibitors towards PL^{pro}. This likely would be done by first generating tight binding UbV(s) towards PL^{pro}, which has successfully been done for the Middle East respiratory syndrome-related coronavirus PL^{pro} in the past (246). Once a UbV is developed towards the enzyme, displacement screens on the PL^{pro}:UbV complex using small molecule inhibitors would be carried out using fluorescence polarization as the displacement readout. The idea is that whichever small molecule inhibitors can displace the UbV from the enzyme must also be binding at the same inhibitory sites and can ultimately act as competitive inhibitors of the enzyme and prevent polyprotein processing, DUB activity and deISGylating activity.

The first step was to recombinantly produce the enzyme as a GST fusion protein, as the phage display library relies on DUBs to be immobilized by their GST tag in order to carry out selections. This also requires that the enzyme be active as a GST fusion protein, ensuring that the active site is accessible with N-term GST fused to the enzyme. We were able to express and purify the protein, which appeared to be active against Ub-AMC (Fig. 4.2A). We provided purified protein to the Sidhu lab for their phage display library selections. They were able to pull down one potent UbV (termed UbV1), which they shared with us via an *E. coli* expression vector. Our main goals were to crystalize the protein complex as Dr. Melnyk's team worked on the displacement screens using small molecule inhibitors.

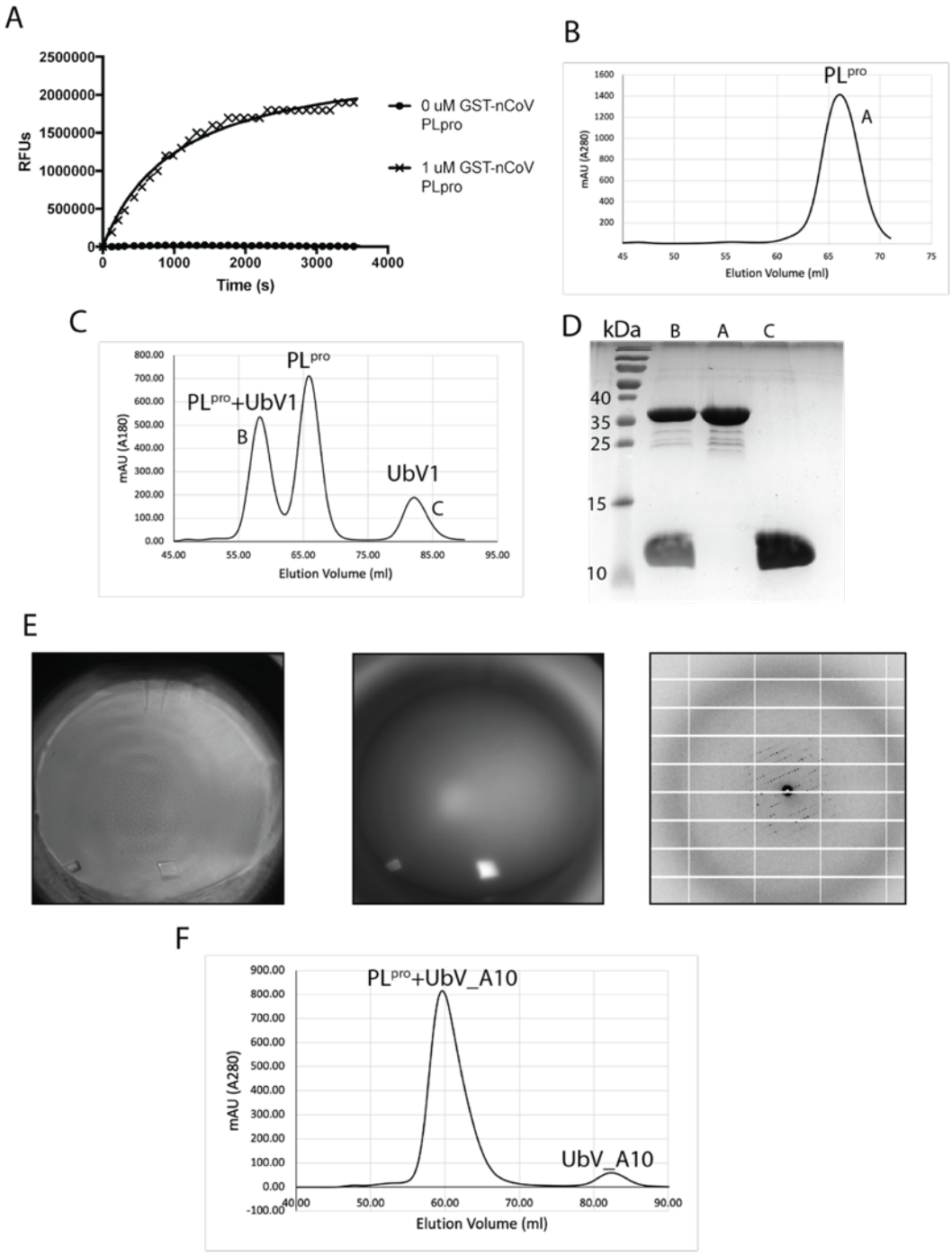


Figure 4.2 SARS-CoV-2 PL^{pro} purification, complexing and crystallization.

A) Ub-AMC assay with GST-tagged PL^{pro}. B) Gel filtration chromatogram of PL^{pro} following GST tag removal. C) Gel filtration chromatogram of PL^{pro} complexing with UbV1. D) SDS-PAGE gel of 3 different peaks from Figs. B and C (PL^{pro} = ~36.0 kDa; UbV1 = ~11.8 kDa). E) Brightfield image of PL^{pro}:UbV1 crystals (left) and fluorescence-based image (middle) and X-ray diffraction of crystals (right). F) Gel filtration chromatogram of PL^{pro} in complex with UbV A10.

4.3.5 PL^{pro} complexing and crystallization

For our studies going forward, we removed the GST tag from PL^{pro} (Fig. 4.2B) for protein complexing. PL^{pro} was a rather challenging protein to express, as a majority of the protein would be found in inclusion bodies; however, we were able to generate enough protein by scaling up the amount of culture used. As a GST-tagged protein, we were only yielding roughly 1.5 mg of protein per liter of culture, so on average 6-8 L of culture was grown for one round of purification to obtain suitable amounts of enzyme for structural studies. UbV1 did complex with PL^{pro}, though there did seem to be an equilibrium between the two proteins (Fig. 4.2C). Even in a fourfold molar excess of UbV1:PL^{pro} there still remained unbound enzyme and un-complexed UbV1, though we were able to separate out the protein complex from free PL^{pro} for crystallization. We were able to crystalize the protein complex in a condition containing 100 mM Bis-TRIS propane (pH 6.5), 200 mM sodium bromide and 20 % PEG 3350 (Fig. 4.2E, left & middle). These crystals were readily reproducible, though never gave strong enough diffraction to collect data from, only weakly diffracting out to ~6-7 Å at best (4.2E, right).

We wanted to switch expression constructs of PL^{pro}, as it was quite laborious to prep milligram amounts of the GST fusion. A construct was generated encoding for PL^{pro} with an N-term 6xHis tag as well as a PreScision protease cleavage site in between the tag and the enzyme, which generated PL^{pro} with the identical amino acid sequence as before (assuming the GST tag is removed). This expression vector was much more efficient, producing 4-5 mg of protein per liter, which made the enzyme preparation more manageable. Eventually, more UbVs were generated by the Sidhu lab that showed more promising binding tendencies. UbV A10 had much better complex forming ability (Fig. 4.2F), binding all PL^{pro} even at only a slight molar excess. Unfortunately, this protein complex failed to crystalize after screening several hundreds of different conditions. This

process is ongoing and hopefully will result in usable crystals. We have also shifted in trying to crystalize the enzyme alone in preparation of receiving small molecule inhibitors for inhibitor-bound PL^{pro} cocrystals. The WT enzyme has also been a challenge to crystalize, which is unsurprising as the enzyme has been mentioned to be challenging to crystalize in its WT or unliganded form (294, 295). These groups have used catalytic Cys mutant protein (Cys111Ser) to generate PL^{pro} Cys111Ser crystals that were subsequently used as seed stock to obtain WT crystals, as they mentioned WT enzyme was recalcitrant in crystallization attempts. We have been able to crystalize PL^{pro} Cys111Ser, which also efficiently couples to the UbVs, and are actively pursuing this avenue to obtain WT crystals with goals of ligand complexing.

4.4 Conclusion

Here we showed that Benyviral PRO domains likely are DUBs. This would be the first determination of this finding, which is quite exciting as a lot of work regarding these enzymes has only been theorized or based on somewhat ambiguous foundational work. We showed that the type member of this genus, BNYVV and the most closely related virus BSBMV have DUB domains coded for within RNA1 of their genome. It is possible that the two other members of this genus (Burdock mottle virus and Rice stripe necrosis virus) also have a similar domain with shared functionality. The difference in poly-Ub chain recognition between BNYVV and BSBMV is curious and must be explored further. This is different than what was seen with the marafiviral PRO domains, which all showed the ability to hydrolyze K48-linked poly-Ub chains and selective recognition towards K63 chains. Neither BSBMV or BNYVV show activity towards K48 chains, indicating the main function of DUB activity is not in modulating 26S proteasomal degradation. The ability of BNYVV to exclusively hydrolyze K63 chains indicates that there is an alternate role of DUB activity, a field that has not been studied with great detail as it pertains to plant viral DUB poly-Ub-K63 recognition. The differences in the amino acid sequence of Ub from beets compared to the human sequence-derived substrates used could also be impactful in the seemingly weak hydrolysis determined. These preliminary results are interesting and should be further investigated.

SARS-CoV-2 PL^{pro} is an intriguing target for antiviral therapeutics as it is essential in polyprotein processing as well as host innate immune response antagonism (268, 296). As soon as the emergence of SARS-CoV-2, PL^{pro} and potential inhibitors were feverishly studied. It is still important to study these enzymes and methods to determine potent inhibitors to fully expand the current antiviral arsenal despite the availability of highly efficacious vaccine therapies. This work being done in collaboration with two University of Toronto labs is still in its infancy, as many

hurdles had to have been overcome, though the preliminary work and findings show promise. The crystallization aspect of this project has been challenging, however, active methods to circumnavigate these challenges are being pursued. A crystal structure of PL^{pro} bound to a UbV would be telling to elucidate which key residues of the UbV impart a much tighter fit in comparison to WT Ub, for which a crystal structure already exists. Ultimately novel small molecule inhibitors will be determined for the enzyme that could prove beneficial as an antiviral therapeutic or at least open the doors for a new class of inhibitor(s) for future coronavirus outbreaks.

As for SARS-CoV-2 PL^{pro}, the main focus remains in crystallization. Current directions towards active site mutant C111S has proven more promising in crystallization for the enzyme alone. This mutant PL^{pro} likely should still be amenable to bind its most potent UbV, albeit potentially at a lower affinity; however, due to the nature of the large number of contacts that typically are involved with Ub(V):DUB complex formation, should still be amenable towards forming a complex with the mutant enzyme and could result in better quality X-ray diffraction.

4.5 Future directions

There still remains a great deal of scope for the benyviral studies. Firstly, to make active site mutations of the PRO domains would be informative to confirm the correct theorized active site residues based on sequence alignments to MRFV PRO. It could be possible that other Cys or His residues may be involved in forming the active site as the overall sequence similarity to marafiviral proteases are quite low. Comparing activities in respect to Ub-AMC hydrolysis to active site mutants would be telling and important to do. Similar poly-Ub chain hydrolysis assays can also be carried out with BNYVV PRO using K63-linked chains to assess active site mutations on poly-Ub chain breakdown. It would be wise to create a truncated form of BSBMV PRO similar to BNYVV PRO, as well as a C-term extended form of BNYVV PRO similar to BSBMV PRO. These forms of the protein could again be used in Ub-AMC assays to assess for heightened or lessened activity. The truncated form of BSBMV PRO could also be used in chain hydrolysis assays to confirm that the extension of BSBMV PRO does not inhibit K63-linked Ub chain hydrolysis.

Given the weak activity towards K63-linked poly Ub chains by BNYVV PRO, alternate Ub chain substrates should be tried, which include K6 (although not present in beet-Ub), K11, K27, K29 and K33-linked chains. This could act to confirm if the weak activity is likely due to their being an alternate linkage form more readily recognized by these proteases. It would also be highly advantageous to have beet-Ub-AMC synthesized and monitor and compare activity towards this substrate in contrast to human Ub-AMC, which would more clearly confirm if weak Ub substrate hydrolysis could merely be attributed to the differences in the Ub sequences.

BNYVV PRO has been attempted in crystallization trials, however, they were unsuccessful. As was the case with OBDV PRO, it was more amenable to crystallizing once

covalently linked to Ub. A pTXB1 DNA construct encoding for beet-Ub instead of human Ub (as done with MRFV PRO-Ub) could be constructed for covalent coupling to the enzyme as carried out with MRFV and OBDV PRO. This structure would be quite interesting as its closest structural homologues determined by Phyre2 are believe by nairovirus OTU domains. Lastly, presenting benyviral DUBs with the UbV phage display library would be of interest to see if any variants could be pulled down. The library is designed against human Ub as a baseline, so it's not certain how benyviral PRO domains would fare in the selection process but would be worth exploring.

References

1. Fisher, S. (2010) Are RNA viruses vestiges of an RNA world? *J. Gen. Philos. Sci.* 10.1007/s10838-010-9119-8
2. Higgs, P. G., and Lehman, N. (2015) The RNA World: Molecular cooperation at the origins of life. *Nat. Rev. Genet.* 10.1038/nrg3841
3. Campillo-Balderas, J. A., Lazcano, A., and Becerra, A. (2015) Viral genome size distribution does not correlate with the antiquity of the host lineages. *Front. Ecol. Evol.* 10.3389/fevo.2015.00143
4. Knipe, D. M., and Howley, P. (2013) *Fields virology: Sixth edition*
5. Krupovic, M., Cvirkaite-Krupovic, V., Iranzo, J., Prangishvili, D., and Koonin, E. V. (2018) Viruses of archaea: Structural, functional, environmental and evolutionary genomics. *Virus Res.* 10.1016/j.virusres.2017.11.025
6. Carrasco-Hernandez, R., Jácome, R., Vidal, Y. L., and de León, S. P. (2017) Are RNA viruses candidate agents for the next global pandemic? A review. *ILAR J.* 10.1093/ilar/ilx026
7. Ahlquist, P., Noueiry, A. O., Lee, W.-M., Kushner, D. B., and Dye, B. T. (2003) Host Factors in Positive-Strand RNA Virus Genome Replication. *J. Virol.* 10.1128/jvi.77.15.8181-8186.2003
8. Latinne, A., Hu, B., Olival, K. J., Zhu, G., Zhang, L., Li, H., Chmura, A. A., Field, H. E., Zambrana-Torrelio, C., Epstein, J. H., Li, B., Zhang, W., Wang, L. F., Shi, Z. L., and Daszak, P. (2020) Origin and cross-species transmission of bat coronaviruses in China. *Nat. Commun.* 10.1038/s41467-020-17687-3
9. Gorbalenya, A. E., Enjuanes, L., Ziebuhr, J., and Snijder, E. J. (2006) Nidovirales: Evolving the largest RNA virus genome. *Virus Res.* 10.1016/j.virusres.2006.01.017
10. Firth, A. E., and Brierley, I. (2012) Non-canonical translation in RNA viruses. *J. Gen. Virol.* **93**, 1385–1409
11. Kamphuis, I. G., Kalk, K. H., Swarte, M. B. A., and Drenth, J. (1984) Structure of papain refined at 1.65 Å resolution. *J. Mol. Biol.* 10.1016/0022-2836(84)90467-4
12. Drenth, J., Jansonius, J. N., Koekoek, R., and Wolthers, B. G. (1971) The structure of papain. *Adv. Protein Chem.* 10.1016/S0065-3233(08)60279-X
13. Mann, K. S., and Sanfaçon, H. (2019) Expanding repertoire of plant positive-strand rna virus proteases. *Viruses*. 10.3390/v11010066
14. Cstorer, A., and Ménard, R. (1994) Catalytic mechanism in papain family of cysteine peptidases. *Methods Enzymol.* 10.1016/0076-6879(94)44035-2
15. Gorbalenya, A. E., Koonin, E. V., and Lai, M. M. C. (1991) Putative papain-related thiol proteases of positive-strand RNA viruses Identification of rubi- and aphthovirus proteases and delineation of a novel conserved domain associated with proteases of rubi-, α- and coronaviruses. *FEBS Lett.* **288**, 201–205
16. Komander, D., Clague, M. J., and Urbé, S. (2009) Breaking the chains: Structure and function of the deubiquitinases. *Nat. Rev. Mol. Cell Biol.* **10**, 550–563
17. Hardy, W. R., and Strauss, J. H. (1989) Processing the nonstructural polyproteins of sindbis virus: nonstructural proteinase is in the C-terminal half of nsP2 and functions both in cis and in trans. *J. Virol.* 10.1128/jvi.63.11.4653-4664.1989
18. Nicklin, M. J. H., Toyoda, H., Murray, M. G., and Wimmer, E. (1986) Proteolytic processing in the replication of polio and related viruses. *Bio/Technology*.

10.1038/nbt0186-33
19. Wellink, J., and Van Kammen, A. (1988) Proteases involved in the processing of viral polyproteins Brief Review. *Arch Virol*. **98**, 1–26
20. Lobigs, M. (1993) Flavivirus premembrane protein cleavage and spike heterodimer secretion require the function of the viral proteinase NS3. *Proc. Natl. Acad. Sci. U. S. A.* 10.1073/pnas.90.13.6218
21. Gupta, G., Lim, L., and Song, J. (2015) NMR and MD studies reveal that the isolated dengue NS3 protease is an intrinsically disordered chymotrypsin fold which absolutely requests NS2B for correct folding and functional dynamics. *PLoS One*. 10.1371/journal.pone.0134823
22. Estoppey, D., Lee, C. M., Janoschke, M., Lee, B. H., Wan, K. F., Dong, H., Mathys, P., Filipuzzi, I., Schuhmann, T., Riedl, R., Aust, T., Galuba, O., McAllister, G., Russ, C., Spiess, M., Bouwmeester, T., Bonamy, G. M. C., and Hoepfner, D. (2017) The Natural Product Cavinafungin Selectively Interferes with Zika and Dengue Virus Replication by Inhibition of the Host Signal Peptidase. *Cell Rep.* 10.1016/j.celrep.2017.03.071
23. Rodamilans, B., Shan, H., Pasin, F., and García, J. A. (2018) Plant viral proteases: Beyond the role of peptide cutters. *Front. Plant Sci.* **9**, 1–11
24. Tong, L. (2002) Viral proteases. *Chem. Rev.* **102**, 4609–4626
25. Bailey-Elkin, B. A., van Kasteren, P. B., Snijder, E. J., Kikkert, M., and Mark, B. L. (2014) Viral OTU Deubiquitinases: A Structural and Functional Comparison. *PLoS Pathog.* 10.1371/journal.ppat.1003894
26. Yost, S. A., and Marcotrigiano, J. (2013) Viral precursor polyproteins: Keys of regulation from replication to maturation. *Curr. Opin. Virol.* 10.1016/j.coviro.2013.03.009
27. Guo, B., Lin, J., and Ye, K. (2011) Structure of the autocatalytic cysteine protease domain of potyvirus helper-component proteinase. *J. Biol. Chem.* **286**, 21937–21943
28. Snijder, E. J., Kikkert, M., and Fang, Y. (2013) Arterivirus molecular biology and pathogenesis. *J. Gen. Virol.* **94**, 2141–63
29. Summers, D. F., and Maizel, J. V. (1968) Evidence for large precursor proteins in poliovirus synthesis. *Proc. Natl. Acad. Sci. U. S. A.* **59**, 966–971
30. Holland, J. J., and Kiehn, E. D. (1968) Specific cleavage of viral proteins as steps in the synthesis and maturation of enteroviruses. *Proc. Natl. Acad. Sci. U. S. A.* **60**, 1015–1022
31. Jacobson, M. F., and Baltimore, D. (1968) Polypeptide cleavages in the formation of poliovirus proteins. *Proc. Natl. Acad. Sci. U. S. A.* **61**, 77–84
32. Penman, S., Scherrer, K., Becker, Y., and Darnell, J. E. (1963) POLYRIBOSOMES IN NORMAL AND POLIOVIRUS-INFECTED HELA CELLS AND THEIR RELATIONSHIP TO MESSENGER-RNA. *Proc. Natl. Acad. Sci.* 10.1073/pnas.49.5.654
33. Jacobson, M. F., Asso, J., and Baltimore, D. (1970) Further evidence on the formation of poliovirus proteins. *J. Mol. Biol.* **49**, 657–669
34. Pelham, H. R. B. (1978) Translation of Encephalomyocarditis Virus RNA in vitro Yields an Active Proteolytic Processing Enzyme. *Eur. J. Biochem.* **85**, 457–462
35. Kitamura, N., Semler, B. L., Rothberg, P. G., Larsen, G. R., Adler, C. J., Dorner, A. J., Emini, E. A., Hanecak, R., Lee, J. J., Van Der Werf, S., Anderson, C. W., and Wimmer, E. (1981) Primary structure, gene organization and polypeptide expression of poliovirus RNA. *Nature*. **291**, 547–553
36. Sárkány, Z., and Polgár, L. (2003) The unusual catalytic triad of poliovirus protease 3C. *Biochemistry*. 10.1021/bi027004w

37. Mosimann, S. C., Cherney, M. M., Sia, S., Plotch, S., and James, M. N. G. (1997) Refined X-ray crystallographic structure of the poliovirus 3C gene product. *J. Mol. Biol.* 10.1006/jmbi.1997.1306
38. Igarashi, H., Yoshino, Y., Miyazawa, M., Horie, H., Ohka, S., and Nomoto, A. (2010) 2A Protease Is Not a Prerequisite for Poliovirus Replication. *J. Virol.* **84**, 5947–5957
39. Alvey, J. C., Wyckoff, E. E., Yu, S. F., Lloyd, R., and Ehrenfeld, E. (1991) cis- and trans-cleavage activities of poliovirus 2A protease expressed in Escherichia coli. *J. Virol.* 10.1128/jvi.65.11.6077-6083.1991
40. Lawson, M. A., and Semler, B. L. (1990) Picornavirus protein processing - Enzymes, substrates, and genetic regulation. *Curr. Top. Microbiol. Immunol.* 10.1007/978-3-642-75602-3_3
41. Rupp, J. C., Sokoloski, K. J., Gebhart, N. N., and Hardy, R. W. (2015) Alphavirus RNA synthesis and non-structural protein functions. *J. Gen. Virol.* 10.1099/jgv.0.000249
42. Atkins, G. J. (2013) The Pathogenesis of Alphaviruses. *ISRN Virol.* 10.5402/2013/861912
43. Jose, J., Snyder, J. E., and Kuhn, R. (2009) A structural and functional perspective of alphavirus replication and assembly. *Future Microbiol.* 10.2217/fmb.09.59
44. de Groot, R. J., Hardy, W. R., Shirako, Y., and Strauss, J. H. (1990) Cleavage-site preferences of Sindbis virus polyproteins containing the non-structural proteinase. Evidence for temporal regulation of polyprotein processing in vivo. *EMBO J.*
45. Vasiljeva, L., Valmu, L., Kääriäinen, L., and Merits, A. (2001) Site-specific Protease Activity of the Carboxyl-terminal Domain of Semliki Forest Virus Replicase Protein nsP2. *J. Biol. Chem.* 10.1074/jbc.M104786200
46. Skern, T., Hampölz, B., Guarné, A., Fita, I., Bergmann, E., Petersen, J., and James, M. N. G. (2014) Structure and Function of Picornavirus Proteinases. *Mol. Biol. Picornavirus*. 10.1128/9781555817916.ch17
47. Steinberger, J., and Skern, T. (2014) The leader proteinase of foot-and-mouth disease virus: Structure-function relationships in a proteolytic virulence factor. *Biol. Chem.* **395**, 1179–1185
48. Sweeney, T. R., Roqué-Rosell, N., Birtley, J. R., Leatherbarrow, R. J., and Curry, S. (2007) Structural and Mutagenic Analysis of Foot-and-Mouth Disease Virus 3C Protease Reveals the Role of the β-Ribbon in Proteolysis. *J. Virol.* 10.1128/jvi.01587-06
49. Gorbalenya, A. E., Koonin, E. V., and Lai, M. M. C. (1991) Putative papain-related thiol proteases of positive-strand RNA viruses Identification of rubi- and aphthovirus proteases and delineation of a novel conserved domain associated with proteases of rubi-, α- and coronaviruses. *FEBS Lett.* **288**, 201–205
50. Guarné, A., Tormo, J., Kirchweger, R., Pfistermueller, D., Fita, I., and Skern, T. (1998) Structure of the foot-and-mouth disease virus leader protease: A papain-like fold adapted for self-processing and eIF4G recognition. *EMBO J.* **17**, 7469–7479
51. Liu, Y., Zhu, Z., Zhang, M., and Zheng, H. (2015) Multifunctional roles of leader protein of foot-and-mouth disease viruses in suppressing host antiviral responses. *Vet. Res.* 10.1186/s13567-015-0273-1
52. Guarné, A., Hampoelz, B., Glaser, W., Carpena, X., Tormo, J., Fita, I., and Skern, T. (2000) Structural and biochemical features distinguish the foot-and-mouth disease virus leader proteinase from other papain-like enzymes. *J. Mol. Biol.* 10.1006/jmbi.2000.4115
53. Svitkin, Y. V., Gradi, A., Imataka, H., Morino, S., and Sonenberg, N. (1999) Eukaryotic Initiation Factor 4GII (eIF4GII), but Not eIF4GI, Cleavage Correlates with Inhibition of

Host Cell Protein Synthesis after Human Rhinovirus Infection. *J. Virol.* 10.1128/jvi.73.4.3467-3472.1999
54. Gradi, A., Svitkin, Y. V., Imataka, H., and Sonenberg, N. (1998) Proteolysis of human eukaryotic translation initiation factor eIF4GII, but not EIF4GI, coincides with the shutoff of host protein synthesis after poliovirus infection. *Proc. Natl. Acad. Sci. U. S. A.* 10.1073/pnas.95.19.11089
55. Malka-Mahieu, H., Newman, M., Désaubry, L., Robert, C., and Vagner, S. (2017) Molecular pathways: The eIF4F translation initiation complex-newopportunities for cancer treatment. *Clin. Cancer Res.* 10.1158/1078-0432.CCR-14-2362
56. Valli, A. A., Gallo, A., Rodamilans, B., López-Moya, J. J., and García, J. A. (2018) The HCPro from the Potyviridae family: an enviable multitasking Helper Component that every virus would like to have. *Mol. Plant Pathol.* **19**, 744–763
57. Wolf, Y. I., Kazlauskas, D., Iranzo, J., Lucía-Sanz, A., Kuhn, J. H., Krupovic, M., Dolja, V. V., and Koonin, E. V. (2018) Origins and evolution of the global RNA virome. *MBio*. 10.1128/mBio.02329-18
58. Mann, K. S., and Sanfaçon, H. (2019) Expanding repertoire of plant positive-strand rna virus proteases. *Viruses*. 10.3390/v11010066
59. Shan, H., Pasin, F., Tzanetakis, I. E., Simón-Mateo, C., García, J. A., and Rodamilans, B. (2018) Truncation of a P1 leader proteinase facilitates potyvirus replication in a non-permissive host. *Mol. Plant Pathol.* 10.1111/mpp.12640
60. Carrington, J. C., Freed, D. D., and Sanders, T. C. (1989) Autocatalytic processing of the potyvirus helper component proteinase in Escherichia coli and in vitro. *J. Virol.* **63**, 4459–4463
61. Peng, C.-W., Peremyslov, V. V., Mushegian, A. R., Dawson, W. O., and Dolja, V. V. (2001) Functional Specialization and Evolution of Leader Proteinases in the Family Closteroviridae. *J. Virol.* **75**, 12153–12160
62. Baulcombe, D. (2004) RNA silencing in plants. *Nature*. **431**, 356–363
63. Calil, I. P., and Fontes, E. P. B. (2017) Plant immunity against viruses: Antiviral immune receptors in focus. *Ann. Bot.* **119**, 711–723
64. Muhammad, T., Zhang, F., Zhang, Y., and Liang, Y. (2019) RNA Interference: A Natural Immune System of Plants to Counteract Biotic Stressors. *Cells*. 10.3390/cells8010038
65. Musidlak, O., Nawrot, R., and Goździcka-Józefiak, A. (2017) Which plant proteins are involved in antiviral defense? Review on in vivo and in vitro activities of selected plant proteins against viruses. *Int. J. Mol. Sci.* 10.3390/ijms18112300
66. Zhang, C., Wu, Z., Li, Y., and Wu, J. (2015) Biogenesis, function, and applications of virus-derived small RNAs in plants. *Front. Microbiol.* 10.3389/fmicb.2015.01237
67. Liu, J. Z., Li, F., and Liu, Y. (2017) *Editorial: Plant immunity against viruses*, 10.3389/fmicb.2017.00520
68. Ji, L., and Chen, X. (2012) Regulation of small RNA stability: Methylation and beyond. *Cell Res.* 10.1038/cr.2012.36
69. Ruiz-García, A. B., Candresse, T., Canales, C., Morán, F., de Oliveira, C. M., Bertolini, E., and Olmos, A. (2020) Molecular characterization of the complete coding sequence of olive leaf yellowing-associated virus. *Plants*. 10.3390/plants9101272
70. Agranovsky, A. A., Koonin, E. V., Boyko, V. P., Maiss, E., Frötschl, R., Lunina, N. A., and Atabekov, J. G. (1994) Beet Yellows Closterovirus: Complete Genome Structure and Identification of a Leader Papain-like Thiol Protease. *Virology*. 10.1006/viro.1994.1034

71. Liu, Y. P., Peremyslov, V. V., Medina, V., and Dolja, V. V. (2009) Tandem leader proteases of Grapevine leafroll-associated virus-2: Host-specific functions in the infection cycle. *Virology*. 10.1016/j.virol.2008.09.035
72. Peng, C.-W., Napuli, A. J., and Dolja, V. V. (2003) Leader Proteinase of Beet Yellows Virus Functions in Long-Distance Transport. *J. Virol.* **77**, 2843–2849
73. Goyal, S. M. (1993) Porcine reproductive and respiratory syndrome. *J. Vet. Diagnostic Investig.* **5**, 656–664
74. Zimmerman, J. J., Dee, S. A., Holtkamp, D. J., Murtaugh, M. P., Stadejek, T., Stevenson, G. W., Torremorell, M., Yang, H., and Zhang, J. (2019) Porcine Reproductive and Respiratory Syndrome Viruses (Porcine Arteriviruses). in *Diseases of Swine*, pp. 685–708, Wiley, 10.1002/9781119350927.ch41
75. Rossow, K. D. (1998) PRRSV Review. *Vet. Pathol*
76. Terpstra, C., Wensvoort, G., and Pol, J. M. (1991) Experimental reproduction of porcine epidemic abortion and respiratory syndrome (mystery swine disease) by infection with Lelystad virus: Koch's postulates fulfilled. *Vet. Q.* **13**, 131–136
77. Zhou, Y. J., Hao, X. F., Tian, Z. J., Tong, G. Z., Yoo, D., An, T. Q., Zhou, T., Li, G. X., Qiu, H. J., Wei, T. C., and Yuan, X. F. (2008) Highly virulent porcine reproductive and respiratory syndrome virus emerged in China. *Transbound. Emerg. Dis.* **55**, 152–164
78. Zhou, L., and Yang, H. (2010) Porcine reproductive and respiratory syndrome in China. *Virus Res.* **154**, 31–37
79. Pileri, E., and Mateu, E. (2016) Review on the transmission porcine reproductive and respiratory syndrome virus between pigs and farms and impact on vaccination. *Vet. Res.* **47**, 1–13
80. Thomann, B., Rushton, J., Schuepbach-Regula, G., and Nathues, H. (2020) Modeling Economic Effects of Vaccination Against Porcine Reproductive and Respiratory Syndrome: Impact of Vaccination Effectiveness, Vaccine Price, and Vaccination Coverage. *Front. Vet. Sci.* **7**, 1–9
81. Lalonde, C., Provost, C., and Gagnon, C. A. (2020) Whole-genome sequencing of porcine reproductive and respiratory syndrome virus from field clinical samples improves the genomic surveillance of the virus. *J. Clin. Microbiol.* **58**, 1–12
82. Rosendal, T., Dewey, C., Friendship, R., Wootton, S., Young, B., and Poljak, Z. (2014) Spatial and temporal patterns of porcine reproductive and respiratory syndrome virus (PRRSV) genotypes in Ontario, Canada, 2004-2007. *BMC Vet. Res.* 10.1186/1746-6148-10-83
83. Renken, C., Nathues, C., Swam, H., Fiebig, K., Weiss, C., Eddicks, M., Ritzmann, M., and Nathues, H. (2021) Application of an economic calculator to determine the cost of porcine reproductive and respiratory syndrome at farm-level in 21 pig herds in Germany. *Porc. Heal. Manag.* **7**, 1–12
84. Montaner-Tarbes, S., del Portillo, H. A., Montoya, M., and Fraile, L. (2019) Key gaps in the knowledge of the porcine respiratory reproductive syndrome virus (PRRSV). *Front. Vet. Sci.* **6**, 1–15
85. Wang, A., Chen, Q., Wang, L., Madson, D., Harmon, K., Gauger, P., Zhang, J., and Li, G. (2019) Recombination between vaccine and field strains of porcine reproductive and respiratory syndrome virus. *Emerg. Infect. Dis.* **25**, 2335–2337
86. Kappes, M. A., and Faaberg, K. S. (2015) PRRSV structure, replication and recombination: Origin of phenotype and genotype diversity. *Virology*. **479–480**, 475–486

87. Snijder, E. J., Kikkert, M., and Fang, Y. (2013) Arterivirus molecular biology and pathogenesis. *J. Gen. Virol.* **94**, 2141–2163
88. Fang, Y., and Snijder, E. J. (2010) The PRRSV replicase: exploring the multifunctionality of an intriguing set of nonstructural proteins. *Virus Res.* **154**, 61–76
89. Li, Y., Treffers, E. E., Napthine, S., Tas, A., Zhu, L., Sun, Z., Bell, S., Mark, B. L., Veelen, P. A. van, Hemert, M. J. van, Firth, A. E., Brierley, I., Snijder, E. J., and Fang, Y. (2014) Transactivation of programmed ribosomal frameshifting by a viral protein. *Proc. Natl. Acad. Sci.* **111**, E2172–E2181
90. den Boon, J. a, Faaberg, K. S., Meulenberg, J. J., Wassenaar, a L., Plagemann, P. G., Gorbalenya, a E., and Snijder, E. J. (1995) Processing and evolution of the N-terminal region of the arterivirus replicase ORF1a protein: identification of two papainlike cysteine proteases. *J. Virol.* **69**, 4500–4505
91. Tian, X., Lu, G., Gao, F., Peng, H., Feng, Y., Ma, G., Bartlam, M., Tian, K., Yan, J., Hilgenfeld, R., and Gao, G. F. (2009) Structure and Cleavage Specificity of the Chymotrypsin-Like Serine Protease (3CLSP/nsp4) of Porcine Reproductive and Respiratory Syndrome Virus (PRRSV). *J. Mol. Biol.* **392**, 977–993
92. Xue, F., Sun, Y., Yan, L., Zhao, C., Chen, J., Bartlam, M., Li, X., Lou, Z., and Rao, Z. (2010) The crystal structure of porcine reproductive and respiratory syndrome virus nonstructural protein Nsp1beta reveals a novel metal-dependent nuclease. *J. Virol.* **84**, 6461–71
93. Sun, Y., Xue, F., Guo, Y., Ma, M., Hao, N., Zhang, X. C., Lou, Z., Li, X., and Rao, Z. (2009) Crystal Structure of Porcine Reproductive and Respiratory Syndrome Virus Leader Protease Nsp1α. *J. Virol.* **83**, 10931–10940
94. Tijms, M. A. (2001) A zinc finger-containing papain-like protease couples subgenomic mRNA synthesis to genome translation in a positive-stranded RNA virus. *Proc. Natl. Acad. Sci.* **98**, 1889–1894
95. Tijms, M. A., van Dinten, L. C., Gorbalenya, A. E., Snijder, E. J., Zhu, L., Sun, Z., Bell, S., Mark, B. L., Veelen, P. A. van, Hemert, M. J. van, Firth, A. E., Brierley, I., Snijder, E. J., and Fang, Y. (2014) A zinc finger-containing papain-like protease couples subgenomic mRNA synthesis to genome translation in a positive-stranded RNA virus. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 1889–94
96. Shi, X., Zhang, X., Wang, F., Wang, L., Qiao, S., Guo, J., Luo, C., Wan, B., Deng, R., and Zhang, G. (2013) The zinc-finger domain was essential for porcine reproductive and respiratory syndrome virus nonstructural protein-1α to inhibit the production of interferon-β. *J. Interf. Cytokine Res.* **33**, 328–334
97. Kroese, M. V., Zevenhoven-Dobbe, J. C., Bos-de Ruijter, J. N. A., Peeters, B. P. H., Meulenberg, J. J. M., Cornelissen, L. A. H. M., and Snijder, E. J. (2008) The nsp 1 α and nsp 1β papain-like autoproteinases are essential for porcine reproductive and respiratory syndrome virus RNA synthesis. *J. Gen. Virol.* **89**, 494–499
98. Li, Y., Tas, A., Snijder, E. J., and Fang, Y. (2012) Identification of porcine reproductive and respiratory syndrome virus ORF1a-encoded non-structural proteins in virus-infected cells. *J. Gen. Virol.* **93**, 829–839
99. Patel, A., Treffers, E. E., Meier, M., Patel, T. R., Stetefeld, J., Snijder, E. J., and Mark, B. L. (2020) Molecular characterization of the RNA-protein complex directing -2/-1 programmed ribosomal frameshifting during arterivirus replicase expression. *J. Biol. Chem.* 10.1074/jbc.RA120.016105

100. Hong, S., Wei, Y., Lin, S., Huang, J., He, W., Yao, J., Chen, Y., Kang, O., Huang, W., and Wei, Z. (2019) Genetic analysis of porcine productive and respiratory syndrome virus between 2013 and 2014 in Southern parts of China: Identification of several novel strains with amino acid deletions or insertions in nsp2. *BMC Vet. Res.* **15**, 1–11
101. Van Kasteren, P. B., Bailey-Elkin, B. A., James, T. W., Ninaber, D. K., Beugeling, C., Khajehpour, M., Snijder, E. J., Mark, B. L., and Kikkert, M. (2013) Deubiquitinase function of arterivirus papain-like protease 2 suppresses the innate immune response in infected host cells. *Proc. Natl. Acad. Sci. U. S. A.* 10.1073/pnas.1218464110
102. Posthuma, C. C., Pedersen, K. W., Lu, Z., Joosten, R. G., Roos, N., Zevenhoven-Dobbe, J. C., and Snijder, E. J. (2008) Formation of the Arterivirus Replication/Transcription Complex: a Key Role for Nonstructural Protein 3 in the Remodeling of Intracellular Membranes. *J. Virol.* **82**, 4480–4491
103. Frias-Staheli, N., Giannakopoulos, N. V., Kikkert, M., Taylor, S. L., Bridgen, A., Paragas, J., Richt, J. A., Rowland, R. R., Schmaljohn, C. S., Lenschow, D. J., Snijder, E. J., García-Sastre, A., and Virgin, H. W. (2007) Ovarian Tumor Domain-Containing Viral Proteases Evade Ubiquitin- and ISG15-Dependent Innate Immune Responses. *Cell Host Microbe*. **2**, 404–416
104. Kappes, M. A., Miller, C. L., and Faaberg, K. S. (2013) Highly Divergent Strains of Porcine Reproductive and Respiratory Syndrome Virus Incorporate Multiple Isoforms of Nonstructural Protein 2 into Virions. *J. Virol.* **87**, 13456–13465
105. Fang, Y., Treffers, E. E., Li, Y., Tas, A., Sun, Z., van der Meer, Y., de Ru, a. H., van Veelen, P. a., Atkins, J. F., Snijder, E. J., and Firth, a. E. (2012) Efficient -2 frameshifting by mammalian ribosomes to synthesize an additional arterivirus protein. *Proc. Natl. Acad. Sci.* **109**, E2920–E2928
106. den Boon, J. A., Faaberg, K. S., Meulenberg, J. J., Wassenaar, A. L., Plagemann, P. G., Gorbalenya, A. E., and Snijder, E. J. (1995) Processing and evolution of the N-terminal region of the arterivirus replicase ORF1a protein: identification of two papainlike cysteine proteases. *J. Virol.* **69**, 4500–5
107. Snijder, E. J., Wassenaar, A. L., and Spaan, W. J. (1992) The 5' end of the equine arteritis virus replicase gene encodes a papainlike cysteine protease. *J. Virol.* 10.1128/jvi.66.12.7040-7048.1992
108. Kroese, M. V., Zevenhoven-Dobbe, J. C., Bos-de Ruijter, J. N. A., Peeters, B. P. H., Meulenberg, J. J. M., Cornelissen, L. A. H. M., and Snijder, E. J. (2008) The nsp1 and nsp1 papain-like autoproteinases are essential for porcine reproductive and respiratory syndrome virus RNA synthesis. *J. Gen. Virol.* **89**, 494–499
109. Han, M., Ke, H., Zhang, Q., and Yoo, D. (2017) Nuclear imprisonment of host cellular mRNA by nsp1β protein of porcine reproductive and respiratory syndrome virus. *Virology*. **505**, 42–55
110. Aravind, L., and Koonin, E. V. (2000) SAP - A putative DNA-binding motif involved in chromosomal organization. *Trends Biochem. Sci.* 10.1016/S0968-0004(99)01537-6
111. de los Santos, T., Diaz-San Segundo, F., Zhu, J., Koster, M., Dias, C. C. A., and Grubman, M. J. (2009) A Conserved Domain in the Leader Proteinase of Foot-and-Mouth Disease Virus Is Required for Proper Subcellular Localization and Function. *J. Virol.* 10.1128/jvi.02112-08
112. Jacks, T., Madhani, H. D., Masiarz, F. R., and Varmus, H. E. (1988) Signals for ribosomal frameshifting in the rous sarcoma virus gag-pol region. *Cell*. **55**, 447–458

113. Jacks, T., and Varmus, H. E. (1985) Expression of the Rous Sarcoma Virus pol Gene by Ribosomal Frameshifting. *Science (80-.).* **230**, 1237–1242
114. Giedroc, D. P., and Cornish, P. V. (2009) Frameshifting RNA pseudoknots: Structure and mechanism. *Virus Res.* **139**, 193–208
115. Dinman, J. D., Ruiz-Echevarria, M. J., and Peltz, S. W. (1998) Translating old drugs into new treatments: Ribosomal frameshifting as a target for antiviral agents. *Trends Biotechnol.* **16**, 190–196
116. Atkins, J. F., Loughran, G., Bhatt, P. R., Firth, A. E., and Baranov, P. V. (2016) Ribosomal frameshifting and transcriptional slippage: From genetic steganography and cryptography to adventitious use. *Nucleic Acids Res.* 10.1093/nar/gkw530
117. Ketteler, R. (2012) On programmed ribosomal frameshifting: The alternative proteomes. *Front. Genet.* **3**, 1–10
118. Miras, M., Allen Miller, W., Truniger, V., and Aranda, M. A. (2017) Non-canonical translation in Plant RNA viruses. *Front. Plant Sci.* 10.3389/fpls.2017.00494
119. den Boon, J. A., Snijder, E. J., Chirnside, E. D., de Vries, A. A., Horzinek, M. C., and Spaan, W. J. (1991) Equine arteritis virus is not a togavirus but belongs to the coronaviruslike superfamily. *J. Virol.* 10.1128/jvi.65.6.2910-2920.1991
120. den Boon, J. A., Snijder, E. J., Chirnside, E. D., de Vries, A. A., Horzinek, M. C., and Spaan, W. J. (1991) Equine arteritis virus is not a togavirus but belongs to the coronaviruslike superfamily. *J. Virol.* **65**, 2910–2920
121. Napthine, S., Ling, R., Finch, L. K., Jones, J. D., Bell, S., Brierley, I., and Firth, A. E. (2017) Protein-directed ribosomal frameshifting temporally regulates gene expression. *Nat. Commun.* **8**, 1–11
122. Li, Y., Treffers, E. E., Napthine, S., Tas, A., Zhu, L., Sun, Z., Bell, S., Mark, B. L., van Veelen, P. a, van Hemert, M. J., Firth, A. E., Brierley, I., Snijder, E. J., and Fang, Y. (2014) Transactivation of programmed ribosomal frameshifting by a viral protein. *Proc. Natl. Acad. Sci. U. S. A.* 10.1073/pnas.1321930111
123. Napthine, S., Treffers, E. E., Bell, S., Goodfellow, I., Fang, Y., Firth, A. E., Snijder, E. J., and Brierley, I. (2016) A novel role for poly(C) binding proteins in programmed ribosomal frameshifting. *Nucleic Acids Res.* **44**, 5491–5503
124. Beura, L. K., Dinh, P. X., Osorio, F. A., and Pattnaik, A. K. (2011) Cellular poly(c) binding proteins 1 and 2 interact with porcine reproductive and respiratory syndrome virus nonstructural protein 1β and support viral replication. *J. Virol.* **85**, 12939–49
125. Wang, L., He, Q., Gao, Y., Guo, X., Ge, X., Zhou, L., and Yang, H. (2012) Interaction of cellular poly(C)-binding protein 2 with nonstructural protein 1β is beneficial to Chinese highly pathogenic porcine reproductive and respiratory syndrome virus replication. *Virus Res.* **169**, 222–230
126. Dickey, T. H., Altschuler, S. E., and Wuttke, D. S. (2013) Single-stranded DNA-binding proteins: Multiple domains for multiple functions. *Structure*. **21**, 1074–1084
127. Makeyev, A. V., and Liebhaber, S. A. (2002) The poly(C)-binding proteins: A multiplicity of functions and a search for mechanisms. *Rna*. **8**, 265–278
128. Makeyev, A. V, and Liebhaber, S. A. (2002) The poly(C)-binding proteins: a multiplicity of functions and a search for mechanisms. *RNA*. **8**, 265–278
129. Choi, H. S., Hwang, C. K., Song, K. Y., Law, P., and Wei, L. (2010) PCBP as transcriptional regulators of gene expression.PDF. **380**, 431–436
130. Hollingworth, D., Candel, A. M., Nicastro, G., Martin, S. R., Briata, P., Gherzi, R., and

Ramos, A. (2012) KH domains with impaired nucleic acid binding as a tool for functional analysis. *Nucleic Acids Res.* 10.1093/nar/gks368
131. Li, Y., Shang, P., Shyu, D., Carrillo, C., Naraghi-Arani, P., Jaing, C. J., Renukaradhya, G. J., Firth, A. E., Snijder, E. J., and Fang, Y. (2018) Nonstructural proteins nsp2TF and nsp2N of porcine reproductive and respiratory syndrome virus (PRRSV) play important roles in suppressing host innate immune responses. *Virology*. **517**, 164–176
132. van Kasteren, P. B., Beugeling, C., Ninaber, D. K., Frias-Staheli, N., van Boheemen, S., Garcia-Sastre, A., Snijder, E. J., and Kikkert, M. (2012) Arterivirus and Nairovirus Ovarian Tumor Domain-Containing Deubiquitinases Target Activated RIG-I To Control Innate Immune Signaling. *J. Virol.* **86**, 773–785
133. Martelli, G. P., Sabanadzovic, S., Sabanadzovic, N. A. G., Edwards, M. C., and Dreher, T. (2002) The family Tymoviridae. *Arch. Virol.* 10.1007/s007050200045
134. International Committee on Taxonomy of Viruses (2020) ICTV Master Species List 2019 v1 (MSL35). *ICTV Master Species List 2019 v1 [Internet].*
135. Nemchinov, L. G., François, S., Roumagnac, P., Ogliastro, M., Hammond, R. W., Mollov, D. S., and Filloux, D. (2018) Characterization of alfalfa virus F, a new member of the genus Marafivirus. *PLoS One*. **13**, 1–11
136. Kim, H., Park, D., and Hahn, Y. (2018) Identification of novel RNA viruses in alfalfa (Medicago sativa): an Alphapartitivirus, a Deltapartitivirus, and a Marafivirus. *Gene*. **638**, 7–12
137. Hogenhout, S. A., Ammar, E.-D., Whitfield, A. E., and Redinbaugh, M. G. (2008) Insect Vector Interactions with Persistently Transmitted Viruses. *Annu. Rev. Phytopathol.* **46**, 327–359
138. Louie, R. (1995) Vascular puncture of maize kernels for the mechanical transmission of maize white line mosaic virus and other viruses of maize. *Phytopathology*. **85**, 139–143
139. Weiland, J. J., and Edwards, M. C. (2011) Linear-motion tattoo machine and prefabricated needle sets for the delivery of plant viruses by vascular puncture inoculation. *Eur. J. Plant Pathol.* **131**, 553–558
140. Dietzgen, R. G., Mann, K. S., and Johnson, K. N. (2016) Plant virus-insect vector interactions: Current and potential future research directions. *Viruses*. 10.3390/v8110303
141. Lefkowitz, E. J., Dempsey, D. M., Hendrickson, R. C., Orton, R. J., Siddell, S. G., and Smith, D. B. (2018) Virus taxonomy: The database of the International Committee on Taxonomy of Viruses (ICTV). *Nucleic Acids Res.* 10.1093/nar/gkx932
142. Invasive species compendium (2021) *CABI*
143. Maccheroni, W., Alegria, M. C., Greggio, C. C., Piazza, P., Kamla, R. F., Zacharias, P. R. A., Bar-joseph, M., Kitajima, E. W., Assumpc, L. C., Camarotte, G., Cardozo, J., Casagrande, E. C., Ferrari, F., Franco, S. F., Giachetto, P. F., Girasol, A., Silva, V. H. A., Souza, L. C. A., Aguilar-vildoso, C. I., Zanca, A. S., Arruda, P., Kitajima, P., Reinach, F. C., Ferro, J. A., and Silva, A. C. R. (2005) Identification and Genomic Characterization of a New Virus (Tymoviridae Family) Associated with Citrus Sudden Death Disease. **79**, 3028–3037
144. Cornejo-Franco, J. F., Alvarez-Quinto, R. A., Grinstead, S., Mollov, D., Karasev, A. V., Ochoa, J., and Quito-Avila, D. F. (2019) A New Tymovirus Isolated From Solanum quitoense: Characterization and Prevalence in Two Solanaceous Crops in Ecuador. *Plant Dis.* **103**, 2246–2251
145. Zambrano, J. L. (2013) Identification of Resistance to Maize rayado fino virus in Maize

Inbred Lines. *Plant Dis.* **97**, 1418–1823
146. Hammond, R. W., and Ramirez, P. (2001) Molecular characterization of the genome of Maize rayado fino virus, the type member of the genus Marafivirus. *Virology*. **282**, 338–347
147. Hunter, M. C., Smith, R. G., Schipanski, M. E., Atwood, L. W., and Mortensen, D. A. (2017) Agriculture in 2050: Recalibrating targets for sustainable intensification. *Bioscience*. **67**, 386–391
148. Bransom, K. L., Weiland, J. J., and Dreher, T. W. (1991) Proteolytic maturation of the 206-kDa nonstructural protein encoded by turnip yellow mosaic virus RNA. *Virology*. **184**, 351–358
149. Chenon, M., Camborde, L., Cheminant, S., and Jupin, I. (2012) A viral deubiquitylating enzyme targets viral RNA-dependent RNA polymerase and affects viral infectivity. *EMBO J.* **31**, 741–753
150. Prod, D., Jakubiec, A., Tournier, V., and Jupin, I. (2003) Targeting of the Turnip Yellow Mosaic Virus 66K Replication Protein to the Chloroplast Envelope Is Mediated by the 140K Protein. *Microbiology*. **77**, 9124–9135
151. Morch, M. D., Boyer, J. C., and Haenni, A. L. (1988) Overlapping open reading frames revealed by complete nucleotide sequencing of turnip yellow mosaic virus genomic RNA. *Nucleic Acids Res.* 10.1093/nar/16.13.6157
152. Edwards, M. C., Weiland, J. J., Todd, J., Stewart, L. R., and Lu, S. (2016) ORF43 of maize rayado fino virus is dispensable for systemic infection of maize and transmission by leafhoppers. *Virus Genes*. **52**, 303–307
153. Edwards, M. C., and Weiland, J. J. (2014) Coat protein expression strategy of oat blue dwarf virus. *Virology*. **450**–**451**, 290–296
154. Izadpanah, K., Ping Zhang, Y., Daubert, S., Masumi, M., and Rowhani, A. (2002) Sequence of the coat protein gene of Bermuda grass etched-line virus, and of the adjacent "marafibox" motif. *Virus Genes*. 10.1023/A:1014516515454
155. Schirawski, J., Voyatzakis, A., Zaccomer, B., Bernardi, F., and Haenni, A.-L. (2000) Identification and Functional Analysis of the Turnip Yellow Mosaic Tymovirus Subgenomic Promoter. *J. Virol.* 10.1128/jvi.74.23.11073-11080.2000
156. Giegé, R., Florentz, C., and Dreher, T. (1993) The TYMV tRNA-like structure. *Biochimie*. 10.1016/0300-9084(93)90063-X
157. Colussi, T. M., Costantino, D. A., Hammond, J. A., Ruehle, G. M., Nix, J. C., and Kieft, J. S. (2014) The structural basis of transfer RNA mimicry and conformational plasticity by a viral RNA. *Nature*. 10.1038/nature13378
158. Morch, M.-D., and Benicourt, C. (1980) Post-Translational Proteolytic Cleavage of In Vitro-Synthesized Turnip Yellow Mosaic Virus RNA-Coded High-Molecular-Weight Proteins. *J. Virol.* **34**, 85–94
159. Morch, M. D., Drugeon, G., Szafranski, P., and Haenni, A. L. (1989) Proteolytic origin of the 150-kilodalton protein encoded by turnip yellow mosaic virus genomic RNA. *J. Virol.* 10.1128/jvi.63.12.5153-5158.1989
160. Bransom, K. L., Wallace, S. E., and Dreher, T. W. (1996) Identification of the cleavage site recognized by the turnip yellow mosaic virus protease. *Virology*. **217**, 404–406
161. Kadare, G., Rozanov, M., and Haenni, A. L. (1995) Expression of the turnip yellow mosaic virus proteinase in Escherichia coli and determination of the cleavage site within the 206 kDa protein. *J. Gen. Virol.* 10.1099/0022-1317-76-11-2853

162. Rozanov, M. N., Drugeon, G., and Haenni, A. L. (1995) Papain-like proteinase of turnip yellow mosaic virus: a prototype of a new viral proteinase group. *Arch. Virol.* 10.1007/BF01309862
163. Jakubiec, A., Drugeon, G., Camborde, L., and Jupin, I. (2007) Proteolytic Processing of Turnip Yellow Mosaic Virus Replication Proteins and Functional Impact on Infectivity. *J. Virol.* **81**, 11402–11412
164. Jakubiec, A., Notaise, J., Tournier, V., Hericourt, F., Block, M. A., Drugeon, G., van Aelst, L., and Jupin, I. (2004) Assembly of Turnip Yellow Mosaic Virus Replication Complexes: Interaction between the Proteinase and Polymerase Domains of the Replication Proteins. *J. Virol.* **78**, 7945–7957
165. Prod'homme, D., Le Panse, S., Drugeon, G., and Jupin, I. (2001) Detection and subcellular localization of the turnip yellow mosaic virus 66K replication protein in infected cells. *Virology*. **281**, 88–101
166. Dreher, T. W. (2004) Turnip yellow mosaic virus: Transfer RNA mimicry, chloroplasts and a C-rich genome. *Mol. Plant Pathol.* 10.1111/J.1364-3703.2004.00236.X
167. Camborde, L., Jupin, I., Planchais, S., Tournier, V., Jakubiec, A., Drugeon, G., Lacassagne, E., Pflieger, S., Chenon, M., and Jupin, I. (2010) The ubiquitin-proteasome system regulates the accumulation of Turnip yellow mosaic virus RNA-dependent RNA polymerase during viral infection. *Plant Cell*. **22**, 3142–3152
168. Komander, D., and Rape, M. (2012) The Ubiquitin Code. *Annu. Rev. Biochem.* **81**, 203–229
169. Bartel, B., and Citovsky, V. (2012) Focus on ubiquitin in plant biology. *Plant Physiol.* **160**, 1
170. Miricescu, A., Goslin, K., and Graciet, E. (2018) Ubiquitylation in plants: Signaling hub for the integration of environmental signals. *J. Exp. Bot.* **69**, 4511–4527
171. Walsh, C. K., and Sadanandom, A. (2014) Ubiquitin chain topology in plant cell signaling: A new facet to an evergreen story. *Front. Plant Sci.* 10.3389/fpls.2014.00122
172. Lombardi, C., Ayach, M., Beaurepaire, L., Chenon, M., Andreani, J., Guerois, R., Jupin, I., and Bressanelli, S. (2013) A Compact Viral Processing Proteinase/Ubiquitin Hydrolase from the OTU Family. *PLoS Pathog.* **9**, e1003560
173. Fieulaine, S., Witte, M. D., Theile, C. S., Ayach, M., Ploegh, H. L., Jupin, I., and Bressanelli, S. (2020) Turnip yellow mosaic virus protease binds ubiquitin suboptimally to fine-tune its deubiquitinase activity . *J. Biol. Chem.* 10.1074/jbc.ra120.014628
174. Somogyi, P., Jenner, A. J., Brierley, I., and Inglis, S. C. (1993) Ribosomal pausing during translation of an RNA pseudoknot. *Mol. Cell. Biol.* **13**, 6931–6940
175. Kontos, H., Napthine, S., and Brierley, I. (2001) Ribosomal Pausing at a Frameshifter RNA Pseudoknot Is Sensitive to Reading Phase but Shows Little Correlation with Frameshift Efficiency. *Mol. Cell. Biol.* 10.1128/mcb.21.24.8657-8670.2001
176. Brierley, I., Digard, P., and Inglis, S. (1989) Characterization of an efficient coronavirus ribosomal frameshifting signal: requirement for an RNA pseudoknot. *Cell*. **57**, 537–547
177. Giedroc, D., and Cornish, P. (2009) Frameshifting RNA Pseudoknots: Structure and Mechanism. *Virus Res.* **139**, 193–208
178. Jacks, T., Townsley, K., Varmus, H. E., and Majors, J. (1987) Two efficient ribosomal frameshifting events are required for synthesis of mouse mammary tumor virus gag-related polyproteins. *Proc. Natl. Acad. Sci. U. S. A.* **84**, 4298–4302
179. Tu, C., Tzeng, T. H., and Bruenn, J. A. (1992) Ribosomal movement impeded at a

pseudoknot required for frameshifting. *Proc. Natl. Acad. Sci. U. S. A.* **89**, 8636–8640
180. Li, Y., Firth, A. E., Brierley, I., Cai, Y., Napthine, S., Wang, T., Yan, X., Kuhn, J. H., and Fang, Y. (2019) Programmed −2/−1 Ribosomal Frameshifting in Simarteriviruses: an Evolutionarily Conserved Mechanism. *J. Virol.* **93**, 1–21
181. Lunney, J. K., Fang, Y., Ladinig, A., Chen, N., Li, Y., Rowland, B., and Renukaradhya, G. J. (2016) Porcine Reproductive and Respiratory Syndrome Virus (PRRSV): Pathogenesis and Interaction with the Immune System. *Annu. Rev. Anim. Biosci.* **4**, 129–154
182. Hu, J., and Zhang, C. (2014) Porcine reproductive and respiratory syndrome virus vaccines: Current status and strategies to a universal vaccine. *Transbound. Emerg. Dis.* 10.1111/tbed.12016
183. Xue, F., Sun, Y., Yan, L., Zhao, C., Chen, J., Bartlam, M., Li, X., Lou, Z., and Rao, Z. (2010) The Crystal Structure of Porcine Reproductive and Respiratory Syndrome Virus Nonstructural Protein Nsp1 Reveals a Novel Metal-Dependent Nuclease. *J. Virol.* **84**, 6461–6471
184. Moore, S. D., and Prevelige, P. E. (2002) A P22 Scaffold Protein Mutation Increases the Robustness of Head Assembly in the Presence of Excess Portal Protein. *J. Virol.* **76**, 10245–10255
185. Snijder, E. J., Wassenaar, A. L., and Spaan, W. J. (1994) Proteolytic processing of the replicase ORF1a protein of equine arteritis virus. *J. Virol.* **68**, 5755–5764
186. Fitzgerald, K. A., McWhirter, S. M., Faia, K. L., Rowe, D. C., Latz, E., Golenbock, D. T., Coyle, A. J., Liao, S. M., and Maniatis, T. (2003) IKKE and TBKI are essential components of the IRF3 signalling pathway. *Nat. Immunol.* **4**, 491–496
187. Seth, R. B., Sun, L., Ea, C. K., and Chen, Z. J. (2005) Identification and characterization of MAVS, a mitochondrial antiviral signaling protein that activates NF-κB and IRF3. *Cell*. **122**, 669–682
188. Bailey-Elkin, B. A., Knaap, R. C. M., Johnson, G. G., Dalebout, T. J., Ninaber, D. K., Van Kasteren, P. B., Bredenbeek, P. J., Snijder, E. J., Kikkert, M., and Mark, B. L. (2014) Crystal structure of the middle east respiratory syndrome coronavirus (MERS-CoV) papain-like protease bound to ubiquitin facilitates targeted disruption of deubiquitinating activity to demonstrate its role in innate immune suppression. *J. Biol. Chem.* **289**, 34667–34682
189. Graham, F. L., and Van der EB, A. . (1973) for the Assay Adenovirus of Infectivity of. *Virology*. **467**, 456–467
190. Unpaired two-tailed Student's t test using GraphPad Prism version 8.1.1 for Windows, GraphPad Software, San Diego, California USA, www.graphpad.com
191. Brown, P. H., and Schuck, P. (2006) Macromolecular size-and-shape distributions by sedimentation velocity analytical ultracentrifugation. *Biophys. J.* **90**, 4651–4661
192. Schuck, P. (2000) Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling. *Biophys. J.* 10.1016/S0006-3495(00)76713-0
193. Schuck, P. (1998) Sedimentation analysis of noninteracting and self-associating solutes using numerical solutions to the Lamm equation. *Biophys. J.* 10.1016/S0006-3495(98)74069-X
194. Johnson, M.L. and M. Straume, Comments on the Analysis of Sedimentation Equilibrium Experiments., in Modern Analytical Ultracentrifugation. Emerging Biochemical and

Biophysical Techniques., Schuster T.M. and L. T.M., Editors. 1994, Birkhäuser: Boston
195. Straume, M., J.D. Veldhuis, and M.L. Johnson, Model-independent quantification of measurement error: empirical estimation of discrete variance function profiles based on standard curves. Methods Enzymol, 1994. 240: p. 121-50
196. Bevington, P.R. and D.K. Robinson, Data reduction and error analysis for the physical sciences. 3rd ed. 2003, Boston: McGraw-Hill. xi, 320 p
197. Tucker, H., A. Wright, G. Deubler, B. Bashir, D. B. Hayes, T. M. Laue and J. Philo (2013). Sedimentation Interpretation Program. New Hampshire, USA, University of New Hampshire.
198. Brautigam, C. A. (2015). Calculations and Publication-Quality Illustrations for Analytical Ultracentrifugation Data. Methods Enzymol. 562: 109-133.
199. Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. Computing In Science & Engineering 9(3): 90-95.
200. Kim, D. N., Thiel, B. C., Mrozowich, T., Hennelly, S. P., Hofacker, I. L., Patel, T. R., and Sanbonmatsu, K. Y. (2020) Zinc-finger protein CNBP alters the 3-D structure of lncRNA Braveheart in solution. *Nat. Commun.* 10.1038/s41467-019-13942-4
201. Majdi Yazdi, M., Saran, S., Mrozowich, T., Lehnert, C., Patel, T. R., Sanders, D. A. R., and Palmer, D. R. J. (2020) Asparagine-84, a regulatory allosteric site residue, helps maintain the quaternary structure of Campylobacter jejuni dihydrodipicolinate synthase. *J. Struct. Biol.* 10.1016/j.jsb.2019.107409
202. Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J., and Svergun, D. I. (2003) PRIMUS: A Windows PC-based system for small-angle scattering data analysis. *J. Appl. Crystallogr.* 10.1107/S0021889803012779
203. Svergun, D. I. (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr.* 10.1107/S0021889892001663
204. Svergun, D. I. (1999) Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys. J.* 10.1016/S0006-3495(99)77443-6
205. Volkov, V. V., and Svergun, D. I. (2003) Uniqueness of ab initio shape determination in small-angle scattering. *J. Appl. Crystallogr.* **36**, 860–864
206. Du, Z., Fenn, S., Tjhen, R., and James, T. L. (2008) Structure of a construct of a human poly(C)-binding protein containing the first and second KH domains reveals insights into its regulatory mechanisms. *J. Biol. Chem.* **283**, 28757–28766
207. Du, Z., Lee, J. K., Fenn, S., Tjhen, R., Stroud, R. M., and James, T. L. (2007) X-ray crystallographic and NMR studies of protein-protein and protein-nucleic acid interactions involving the KH domains from human poly(C)-binding protein-2. *Rna*. **13**, 1043–1051
208. Li, S., Olson, W. K., and Lu, X. J. (2019) Web 3DNA 2.0 for the analysis, visualization, and modeling of 3D nucleic acid structures. *Nucleic Acids Res.* **47**, W26–W34
209. The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC
210. Fenn, S., Du, Z., Lee, J. K., Tjhen, R., Stroud, R. M., and James, T. L. (2007) Crystal structure of the third KH domain of human poly (C) -binding protein-2 in complex with a C-rich ˚ resolution strand of human telomeric DNA at 1 . 6 A. *Nucleic Acids Res.* **35**, 2651–2660
211. Du, Z., Lee, J. K., Tjhen, R., Li, S., Pan, H., Stroud, R. M., and James, T. L. (2005) Crystal structure of the first KH domain of human poly(C)-binding protein-2 in complex with a C-rich strand of human telomeric DNA at 1.7 Å. *J. Biol. Chem.* **280**, 38823–38830

212. Huynh, K., and Partch, C. L. (2016) Current Protocols in Protein Science: Analysis of protein stability and ligand interactions by thermal shift assay. *Curr Protoc Protein Sci*
213. Tomonaga, T., and Levens, D. (1995) Heterogeneous nuclear ribonucleoprotein K is a DNA-binding transactivator. *J. Biol. Chem.* **270**, 4875–4881
214. Yoga, Y. M. K., Traore, D. A. K., Sidiqi, M., Szeto, C., Pendini, N. R., Barker, A., Leedman, P. J., Wilce, J. A., and Wilce, M. C. J. (2012) Contribution of the first K-homology domain of poly(C)-binding protein 1 to its affinity and specificity for C-rich oligonucleotides. *Nucleic Acids Res.* 10.1093/nar/gks058
215. Li, Y., Zhu, L., Lawson, S. R., and Fang, Y. (2013) Targeted mutations in a highly conserved motif of the nsp1 β protein impair the interferon antagonizing activity of porcine reproductive and respiratory syndrome virus. *J. Gen. Virol.* **94**, 1972–1983
216. Beura, L. K., Sarkar, S. N., Kwon, B., Subramaniam, S., Jones, C., Pattnaik, A. K., and Osorio, F. A. (2010) Porcine Reproductive and Respiratory Syndrome Virus Nonstructural Protein 1 Modulates Host Innate Immune Response by Antagonizing IRF3 Activation. *J. Virol.* **84**, 1574–1584
217. Sun, Z., Chen, Z., Lawson, S. R., and Fang, Y. (2010) The Cysteine Protease Domain of Porcine Reproductive and Respiratory Syndrome Virus Nonstructural Protein 2 Possesses Deubiquitinating and Interferon Antagonism Functions. *J. Virol.* **84**, 7832–7846
218. Beura, L., Dinh, P., Osorio, F., and Pattnaik, A. Cellular poly(c) binding proteins 1 and 2 interact with porcine reproductive and respiratory syndrome virus nonstructural protein 1β and support viral replication. *J Virol*. 10.1128/jvi.05177-11
219. Chen, Z., Lawson, S., Sun, Z., Zhou, X., Guan, X., Christopher-Hennings, J., Nelson, E. A., and Fang, Y. (2010) Identification of two auto-cleavage products of nonstructural protein 1 (nsp1) in porcine reproductive and respiratory syndrome virus infected cells: nsp1 function as interferon antagonist. *Virology*. 10.1016/j.virol.2009.11.033
220. Han, M., Kim, C. Y., Rowland, R. R. R., Fang, Y., Kim, D., and Yoo, D. (2014) Biogenesis of non-structural protein 1 (nsp1) and nsp1-mediated type I interferon modulation in arteriviruses. *Virology*. 10.1016/j.virol.2014.04.028
221. Li, Y., Shyu, D.-L., Shang, P., Bai, J., Ouyang, K., Dhakal, S., Hiremath, J., Binjawadagi, B., Renukaradhya, G. J., and Fang, Y. (2016) Mutations in a Highly Conserved Motif of nsp1β Protein Attenuate the Innate Immune Suppression Function of Porcine Reproductive and Respiratory Syndrome Virus. *J. Virol.* 10.1128/jvi.03069-15
222. Schuck, P. (1998) Sedimentation analysis of noninteracting and self-associating solutes using numerical solutions to the Lamm equation. *Biophys. J.* **75**, 1503–1512
223. Madeira, F., Park, Y. M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P., Tivey, A. R. N., Potter, S. C., Finn, R. D., and Lopez, R. (2019) The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* **47**, W636–W641
224. E. Krissinel and K. Henrick (2007). Inference of macromolecular assemblies from crystalline state. J. Mol. Biol. 372, 774-797.
225. Guinier, A., Fournet, G., Walker, C. B., and Vineyard, G. H. (1956) Small-Angle Scattering of X-Rays . *Phys. Today*. 10.1063/1.3060069
226. Pérez, J., and Vachette, P. (2017) A successful combination: Coupling SE-HPLC with SAXS. in *Advances in Experimental Medicine and Biology*, 10.1007/978-981-10-6038-0_11
227. Patel, T. R., Chojnowski, G., Astha, Koul, A., McKenna, S. A., and Bujnicki, J. M. (2017) Structural studies of RNA-protein complexes: A hybrid approach involving

hydrodynamics, scattering, and computational methods. *Methods*. 10.1016/j.ymeth.2016.12.002
228. Durand, D., Vivès, C., Cannella, D., Pérez, J., Pebay-Peyroula, E., Vachette, P., and Fieschi, F. (2010) NADPH oxidase activator p67phox behaves in solution as a multidomain protein with semi-flexible linkers. *J. Struct. Biol.* 10.1016/j.jsb.2009.08.009
229. Patel, T. R., Bernards, C., Meier, M., McEleney, K., Winzor, D. J., Koch, M., and Stetefeld, J. (2014) Structural elucidation of full-length nidogen and the laminin-nidogen complex in solution. *Matrix Biol.* 10.1016/j.matbio.2013.07.009
230. Mrozowich, T., Henrickson, A., Demeler, B., and Patel, T. R. (2020) Nanoscale structure determination of Murray Valley encephalitis and Powassan virus non-coding RNAs. *Viruses*. 10.3390/v12020190
231. Ortega, A., Amorós, D., and García De La Torre, J. (2011) Prediction of hydrodynamic and other solution properties of rigid proteins from atomic- and residue-level models. *Biophys. J.* 10.1016/j.bpj.2011.06.046
232. Meier, M., Moya-Torres, A., Krahn, N. J., McDougall, M. D., Orriss, G. L., McRae, E. K. S., Booy, E. P., McEleney, K., Patel, T. R., McKenna, S. A., and Stetefeld, J. (2018) Structure and hydrodynamics of a DNA G-quadruplex with a cytosine bulge. *Nucleic Acids Res.* 10.1093/nar/gky307
233. Zhou, Z. H. (2011) Atomic resolution cryo electron microscopy of macromolecular complexes. in *Advances in Protein Chemistry and Structural Biology*, 10.1016/B978-0-12-386507-6.00001-4
234. Yuan, S., Peng, L., Park, J. J., Hu, Y., Devarkar, S. C., Dong, M. B., Shen, Q., Wu, S., Chen, S., Lomakin, I. B., and Xiong, Y. (2020) Nonstructural Protein 1 of SARS-CoV-2 Is a Potent Pathogenicity Factor Redirecting Host Protein Synthesis Machinery toward Viral RNA. *Mol. Cell*. 10.1016/j.molcel.2020.10.034
235. Bartenschlager, R., Ahlborn-Laake, L., Mous, J., and Jacobsen, H. (1994) Kinetic and structural analyses of hepatitis C virus polyprotein processing. *J. Virol.* **68**, 5045–55
236. Dolja, V. V., and Koonin, E. V. (2011) Common origins and host-dependent diversity of plant and animal viromes. *Curr. Opin. Virol.* 10.1016/j.coviro.2011.09.007
237. Ruiz-García, A. B., Sabaté, J., Lloria, O., Laviña, A., Batlle, A., and Olmos, A. (2017) First report of grapevine syrah virus-1 in grapevine in Spain. *Plant Dis.* 10.1094/PDIS-05-17-0700-PDN
238. Hammond, R. W., Edwards, M. C., and Ramirez, P. (2011) *The Springer Index of Viruses* (Tidona, C., and Darai, G. eds), Springer New York, New York, NY, 10.1007/978-0-387-95919-1
239. Nault, L. R. (1980) Leafhopper Transmission and Host Range of Maize Rayado Fino Virus. *Phytopathology*. **70**, 709
240. Edwards, M. C., Weiland, J. J., Todd, J., and Stewart, L. R. (2015) Infectious Maize rayado fino virus from Cloned cDNA. **105**, 833–839
241. Moriceau, L., Jomat, L., Bressanelli, S., Alcaide-Loridan, C., and Jupin, I. (2017) Identification and molecular characterization of the chloroplast targeting domain of turnip yellow mosaic virus replication proteins. *Front. Plant Sci.* **8**, 1–19
242. Liu, H., and Naismith, J. H. (2008) An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *BMC Biotechnol.* 10.1186/1472-6750-8-91
243. Borodovsky, A., Ovaa, H., Kolli, N., Gan-Erdene, T., Wilkinson, K. D., Ploegh, H. L., and

Kessler, B. M. (2002) Chemistry-based functional proteomics reveals novel members of the deubiquitinating enzyme family. *Chem. Biol.* **9**, 1149–1159
244. Messick, T. E., Russell, N. S., Iwata, A. J., Sarachan, K. L., Shiekhattar, R., Shanks, J. R., Reyes-Turcu, F. E., Wilkinson, K. D., and Marmorstein, R. (2008) Structural basis for ubiquitin recognition by the Otu1 ovarian tumor domain protein. *J. Biol. Chem.* **283**, 11038–11049
245. Ernst, A., Avvakumov, G., Tong, J., Fan, Y., Zhao, Y., Alberts, P., Persaud, A., Walker, J. R., Neculai, A. M., Neculai, D., Vorobyov, A., Garg, P., Beatty, L., Chan, P. K., Juang, Y. C., Landry, M. C., Yeh, C., Zeqiraj, E., Karamboulas, K., Allali-Hassani, A., Vedadi, M., Tyers, M., Moffat, J., Sicheri, F., Pelletier, L., Durocher, D., Raught, B., Rotin, D., Yang, J., Moran, M. F., Dhe-Paganon, S., and Sidhu, S. S. (2013) A strategy for modulation of enzymes in the ubiquitin system. *Science (80-.).* 10.1126/science.1230161
246. Zhang, W., Bailey-elkin, B. A., Knaap, R. C. M., Khare, B., Dalebout, J., Johnson, G. G., Kasteren, P. B. Van, Mcleish, N. J., Gu, J., He, W., Kikkert, M., Mark, B. L., and Sidhu, S. S. (2017) Potent and selective inhibition of pathogenic viruses by engineered ubiquitin variants. 10.1371/journal.ppat.1006372
247. Yoo, S. D., Cho, Y. H., and Sheen, J. (2007) Arabidopsis mesophyll protoplasts: A versatile cell system for transient gene expression analysis. *Nat. Protoc.* **2**, 1565–1572
248. Wu, F. H., Shen, S. C., Lee, L. Y., Lee, S. H., Chan, M. T., and Lin, C. S. (2009) Tape-arabidopsis sandwich - A simpler arabidopsis protoplast isolation method. *Plant Methods*. **5**, 1–10
249. Kabsch, W. (2010) XDS. *Acta Crystallogr. - Sect. D Biol. Crystallogr.*
250. Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R., and Leslie, A. G. W. (2011) iMOSFLM: A new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67**, 271–281
251. Evans, P. R., and Murshudov, G. N. (2013) How good are my data and what is the resolution? *Acta Crystallogr. Sect. D Biol. Crystallogr.* 10.1107/S0907444913000061
252. Potterton, L., Agirre, J., Ballard, C., Cowtan, K., Dodson, E., Evans, P. R., Jenkins, H. T., Keegan, R., Krissinel, E., Stevenson, K., Lebedev, A., McNicholas, S. J., Nicholls, R. A., Noble, M., Pannu, N. S., Roth, C., Sheldrick, G., Skubak, P., Turkenburg, J., Uski, V., Von Delft, F., Waterman, D., Wilson, K., Winn, M., and Wojdyr, M. (2018) CCP 4 i 2: The new graphical user interface to the CCP 4 program suite. *Acta Crystallogr. Sect. D Struct. Biol.* 10.1107/S2059798317016035
253. Liebschner, D., Afonine, P. V., Baker, M. L., Bunkoczi, G., Chen, V. B., Croll, T. I., Hintze, B., Hung, L. W., Jain, S., McCoy, A. J., Moriarty, N. W., Oeffner, R. D., Poon, B. K., Prisant, M. G., Read, R. J., Richardson, J. S., Richardson, D. C., Sammito, M. D., Sobolev, O. V., Stockwell, D. H., Terwilliger, T. C., Urzhumtsev, A. G., Videau, L. L., Williams, C. J., and Adams, P. D. (2019) Macromolecular structure determination using X-rays, neutrons and electrons: Recent developments in Phenix. *Acta Crystallogr. Sect. D Struct. Biol.* 10.1107/S2059798319011471
254. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 10.1107/S0907444910007493
255. Bransom, K. L., Wallace, S. E., and Dreher, T. W. (1996) Identification of the cleavage site recognized by the turnip yellow mosaic virus protease. *Virology*. 10.1006/viro.1996.0131
256. Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R.,

McWilliam, H., Remmert, M., Söding, J., Thompson, J. D., and Higgins, D. G. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* 10.1038/msb.2011.75
257. Hua, Z., and Vierstra, R. D. (2011) The Cullin-RING Ubiquitin-Protein Ligases. *Annu. Rev. Plant Biol.* **62**, 299–334
258. Xu, P., and Peng, J. (2006) Dissecting the ubiquitin pathway by mass spectrometry. *Biochim. Biophys. Acta - Proteins Proteomics*. 10.1016/j.bbapap.2006.09.004
259. Johnson, A., and Vert, G. (2016) Unraveling K63 polyubiquitination networks by sensor-based proteomics. *Plant Physiol.* **171**, 1808–1820
260. Braten, O., Livneh, I., Ziv, T., Admon, A., Kehat, I., Caspi, L. H., Gonen, H., Bercovich, B., Godzik, A., Jahandideh, S., Jaroszewski, L., Sommer, T., Kwon, Y. T., Guharoy, M., Tompa, P., and Ciechanover, A. (2016) Numerous proteins with unique characteristics are degraded by the 26S proteasome following monoubiquitination. *Proc. Natl. Acad. Sci. U. S. A.* 10.1073/pnas.1608644113
261. Saeki, Y., Kudo, T., Sone, T., Kikuchi, Y., Yokosawa, H., Toh-e, A., and Tanaka, K. (2009) Lysine 63-linked polyubiquitin chain may serve as a targeting signal for the 26S proteasome. *EMBO J.* 10.1038/emboj.2008.305
262. Zhou, B., and Zeng, L. (2017) Conventional and unconventional ubiquitination in plant immunity. *Mol. Plant Pathol.* **18**, 1313–1330
263. Isono, E., and Nagel, M. K. (2014) Deubiquitylating enzymes and their emerging role in plant biology. *Front. Plant Sci.* **5**, 1–6
264. Bailey-Elkin, B. A., Knaap, R. C. M., Kikkert, M., and Mark, B. L. (2017) Structure and Function of Viral Deubiquitinating Enzymes. *J. Mol. Biol.* **429**, 3441–3470
265. Dang, L. C., Melandri, F. D., and Stein, R. L. (1998) Kinetic and mechanistic studies on the hydrolysis of ubiquitin C- terminal 7-amido-4-methylcoumarin by deubiquitinating enzymes. *Biochemistry*. 10.1021/bi9723360
266. Clementz, M. A., Chen, Z., Banach, B. S., Wang, Y., Sun, L., Ratia, K., Baez-Santos, Y. M., Wang, J., Takayama, J., Ghosh, A. K., Li, K., Mesecar, A. D., and Baker, S. C. (2010) Deubiquitinating and Interferon Antagonism Activities of Coronavirus Papain-Like Proteases. *J. Virol.* 10.1128/jvi.02406-09
267. Balakirev, M. Y., Jaquinod, M., Haas, A. L., and Chroboczek, J. (2002) Deubiquitinating Function of Adenovirus Proteinase. *J. Virol.* 10.1128/jvi.76.12.6323-6331.2002
268. Shin, D., Mukherjee, R., Grewe, D., Bojkova, D., Baek, K., Bhattacharya, A., Schulz, L., Widera, M., Mehdipour, A. R., Tascher, G., Geurink, P. P., Wilhelm, A., van der Heden van Noort, G. J., Ovaa, H., Müller, S., Knobeloch, K. P., Rajalingam, K., Schulman, B. A., Cinatl, J., Hummer, G., Ciesek, S., and Dikic, I. (2020) Papain-like protease regulates SARS-CoV-2 viral spread and innate immunity. *Nature*. 10.1038/s41586-020-2601-5
269. Capodagli, G. C., Deaton, M. K., Baker, E. A., Lumpkin, R. J., and Pegan, S. D. (2013) Diversity of Ubiquitin and ISG15 Specificity among Nairoviruses' Viral Ovarian Tumor Domain Proteases. *J. Virol.* **87**, 3815–3827
270. An, G. (1985) High Efficiency Transformation of Cultured Tobacco Cells. *Plant Physiol.* **79**, 568–570
271. Emmott, E., de Rougemont, A., Hosmillo, M., Lu, J., Fitzmaurice, T., Haas, J., and Goodfellow, I. (2019) Polyprotein processing and intermolecular interactions within the viral replication complex spatially and temporally control norovirus protease activity. *J. Biol. Chem.* 10.1074/jbc.RA118.006780

272. Rausalu, K., Utt, A., Quirin, T., Varghese, F. S., Žusinaite, E., Das, P. K., Ahola, T., and Merits, A. (2016) Chikungunya virus infectivity, RNA replication and non-structural polyprotein processing depend on the nsP2 protease's active site cysteine residue. *Sci. Rep.* 10.1038/srep37124
273. Gildenhuys, S. (2020) Expanding our understanding of the role polyprotein conformation plays in the coronavirus life cycle. *Biochem. J.* 10.1042/BCJ20200223
274. Kadare, G., Rozanov, M., and Haenni, A. L. (1995) Expression of the turnip yellow mosaic virus proteinase in Escherichia coli and determination of the cleavage site within the 206 kDa protein. *J. Gen. Virol.* **76**, 2853–2857
275. Espadaler, J., Querol, E., Aviles, F. X., and Oliva, B. (2006) Identification of function-associated loop motifs and application to protein function prediction. *Bioinformatics*. 10.1093/bioinformatics/btl382
276. Holm, L. (2020) DALI and the persistence of protein shape. *Protein Sci.* 10.1002/pro.3749
277. Jupin, I., Ayach, M., Jomat, L., Fieulaine, S., and Bressanelli, S. (2017) *A mobile loop near the active site acts as a switch between the dual activities of a viral protease/deubiquitinase*, 10.1371/journal.ppat.1006714
278. Krissinel, E., and Henrick, K. (2007) Inference of Macromolecular Assemblies from Crystalline State. *J. Mol. Biol.* 10.1016/j.jmb.2007.05.022
279. Dikic, I., Wakatsuki, S., and Walters, K. J. (2009) Ubiquitin-binding domains from structures to functions. *Nat. Rev. Mol. Cell Biol.* 10.1038/nrm2767
280. Fieulaine, S., Witte, M. D., Theile, C. S., Ayach, M., Ploegh, H. L., Jupin, I., and Bressanelli, S. (2020) Turnip yellow mosaic virus protease binds ubiquitin suboptimally to fine-tune its deubiquitinase activity. *J. Biol. Chem.* **295**, 13769–13783
281. Gilmer, D., and Ratti, C. (2017) ICTV virus taxonomy profile: Benyviridae. *J. Gen. Virol.* **98**, 1571–1572
282. Rush, C. M. (2003) Ecology and Epidemiology of Benyviruses and Plasmodiophorid Vectors. *Annu. Rev. Phytopathol.* **41**, 567–592
283. Chiba, S., Kondo, H., Miyanishi, M., Andika, I. B., Han, C., and Tamada, T. (2011) The evolutionary history of beet necrotic yellow vein virus deduced from genetic variation, geographical origin and spread, and the breaking of host resistance. *Mol. Plant-Microbe Interact.* **24**, 207–218
284. Ziegler, V., Richards, K., Guilley, H., Jonard, G., and Putz, C. (1985) Cell-free Translation of Beet Necrotic Yellow Vein Virus: Readthrough of the Coat Protein Cistron. *J. Gen. Virol.* **66**, 2079–2087
285. Naqvi, A. A. T., Fatima, K., Mohammad, T., Fatima, U., Singh, I. K., Singh, A., Atif, S. M., Hariprasad, G., Hasan, G. M., and Hassan, M. I. (2020) Insights into SARS-CoV-2 genome, structure, evolution, pathogenesis and therapies: Structural genomics approach. *BBA - Mol. Basis Dis.*
286. Wu, C., Liu, Y., Yang, Y., Zhang, P., Zhong, W., Wang, Y., Wang, Q., Xu, Y., Li, M., Li, X., Zheng, M., Chen, L., and Li, H. (2020) Analysis of therapeutic targets for SARS-CoV-2 and discovery of potential drugs by computational methods. *Acta Pharm. Sin. B*. 10.1016/j.apsb.2020.02.008
287. Barretto, N., Jukneliene, D., Ratia, K., Chen, Z., Mesecar, A. D., and Baker, S. C. (2005) The Papain-Like Protease of Severe Acute Respiratory Syndrome Coronavirus Has Deubiquitinating Activity. *J. Virol.* **79**, 15189–15198
288. Mielech, A. M., Kilianski, A., Baez-Santos, Y. M., Mesecar, A. D., and Baker, S. C.

(2014) MERS-CoV papain-like protease has deISGylating and deubiquitinating activities. *Virology*. **450**–**451**, 64–70
289. Békés, M., van der Heden van Noort, G. J., Ekkebus, R., Ovaa, H., Huang, T. T., and Lima, C. D. (2016) Recognition of Lys48-Linked Di-ubiquitin and Deubiquitinating Activities of the SARS Coronavirus Papain-like Protease. *Mol. Cell*. **62**, 572–585
290. Narasimhan, J., Wang, M., Fu, Z., Klein, J. M., Haas, A. L., and Kim, J. J. P. (2005) Crystal structure of the interferon-induced ubiquitin-like protein ISG15. *J. Biol. Chem.* 10.1074/jbc.M502814200
291. Malakhov, M. P., Malakhova, O. A., Il Kim, K., Ritchie, K. J., and Zhang, D. E. (2002) UBP43 (USP18) specifically removes ISG15 from conjugated proteins. *J. Biol. Chem.* 10.1074/jbc.M109078200
292. Forni, G., Mantovani, A., Forni, G., Mantovani, A., Moretta, L., Rappuoli, R., Rezza, G., Bagnasco, A., Barsacchi, G., Bussolati, G., Cacciari, M., Cappuccinelli, P., Cheli, E., Guarini, R., Bacci, M. L., Mancini, M., Marcuzzo, C., Morrone, M. C., Parisi, G., Pasquino, G., Patrono, C., Curzio, A. Q., Remuzzi, G., Roncaglia, A., Schiaffino, S., and Vineis, P. (2021) COVID-19 vaccines: where we stand and challenges ahead. *Cell Death Differ.* 10.1038/s41418-020-00720-9
293. Kouznetsova, V. L., Zhang, A., Tatineni, M., Miller, M. A., and Tsigelny, I. F. (2020) Potential COVID-19 papain-like protease PLpro inhibitors: Repurposing FDA-approved drugs. *PeerJ*. **8**, 1–13
294. Osipiuk, J., Azizi, S. A., Dvorkin, S., Endres, M., Jedrzejczak, R., Jones, K. A., Kang, S., Kathayat, R. S., Kim, Y., Lisnyak, V. G., Maki, S. L., Nicolaescu, V., Taylor, C. A., Tesar, C., Zhang, Y. A., Zhou, Z., Randall, G., Michalska, K., Snyder, S. A., Dickinson, B. C., and Joachimiak, A. (2021) Structure of papain-like protease from SARS-CoV-2 and its complexes with non-covalent inhibitors. *Nat. Commun.* **12**, 1–9
295. Fu, Z., Huang, B., Tang, J., Liu, S., Liu, M., Ye, Y., Liu, Z., Xiong, Y., Zhu, W., Cao, D., Li, J., Niu, X., Zhou, H., Zhao, Y. J., Zhang, G., and Huang, H. (2021) The complex structure of GRL0617 and SARS-CoV-2 PLpro reveals a hot spot for antiviral drug discovery. *Nat. Commun.* **12**, 1–12
296. Klemm, T., Ebert, G., Calleja, D. J., Allison, C. C., Richardson, L. W., Bernardini, J. P., Lu, B. G., Kuchel, N. W., Grohmann, C., Shibata, Y., Gan, Z. Y., Cooney, J. P., Doerflinger, M., Au, A. E., Blackmore, T. R., Heden van Noort, G. J., Geurink, P. P., Ovaa, H., Newman, J., Riboldi-Tunnicliffe, A., Czabotar, P. E., Mitchell, J. P., Feltham, R., Lechtenberg, B. C., Lowes, K. N., Dewson, G., Pellegrini, M., Lessene, G., and Komander, D. (2020) Mechanism and inhibition of the papain-like protease, PLpro, of SARS-CoV-2. *EMBO J.* 10.15252/embj.2020106275
297. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., and Sternberg, M. J. E. (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* 10.1038/nprot.2015.053
298. Breeze, E. (2020) Ubiquitous Ubiquitin: The K63 Ubiquitinome. *Plant Cell*. 10.1105/tpc.19.00898
299. Tracz, M., and Bialek, W. (2021) Beyond K48 and K63: non-canonical protein ubiquitination. *Cell. Mol. Biol. Lett.* 10.1186/s11658-020-00245-6