RICE UNIVERSITY

By

Yu Ouyang

A THESIS SUBMITTED
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE

Doctor of Philosophy

APPROVED, THESIS COMMITTEE

Yizhi Tao (Mar 29, 2023 09:18 CDT)

Yizhi Jane Tao

Marcos deMoraes (Mar 29, 2023 10:33 CDT)

Marcos de Moraes

Michael C. Gustin

Isaac Hilton

HOUSTON, TEXAS

March 2023

ABSTRACT

"Expression and Assembly of Human Picobirnavirus (hPBV) Using a Plasmid-Based Expression System"

by

Yu Ouyang

Picobirnavirus (PBV) is a small (35 nm in diameter, i.e., "pico"), non-enveloped, bi-segmented dsRNA (i.e., "bi-RNA") virus. Since the discovery of PBV in 1988, PBV has been detected in various host species, including humans, mammals, birds, reptiles, and even in environments like sewage. The actual host of PBV remains controversial, and no infection model has been established to unveil the pathology or pathogenesis associated with PBV infection. In this study, I constructed a plasmid-based expression system for hPBV by expressing the two viral RNA segments in the *E. coli* Rosetta 2 strain. All predicted viral proteins, i.e., ORF1, ORF2, CP, and RdRP, were detected upon the viral RNA expression, indicating efficient translation from inherent ribosomal binding sites. Purified recombinant hPBV capsids were found to comprise: both viral RNA segments, ORF2, and RdRP. Viral RNA segments were preferentially packaged into the capsids over host mRNAs, presumably due to packaging signal (PS) sequences at the terminal ends. Such terminal PS sequences were able to carry a non-viral RNA sequence into the recombinant hPBV. ORF2 protein was observed in the VLPs for the first time. By engineering a series of truncation mutants, it was determined that the N-terminal domain of ORF2 is responsible for its incorporation into the capsids. Through immunoprecipitation assays, it was found that hPBV RdRP directly interacts with CP into assembly intermediates. Still, RdRP was only present at deficient levels in assembled capsids at its overexpression. This work presents a plasmid-based expression system to fully dissect molecular determinants for hPBV assembly and genome packaging. Results from this study indicate that hPBV is indeed a prokaryotic virus, but *E. coli* is likely not the natural host of hPBV. Co-immunoprecipitation of recombinant hPBV with lysates of

different microbiota species in future studies would reveal the identity of the hPBV native host and establish the PBV infection model.

Acknowledgments

I appreciate all help and support I received at Rice. First and foremost, I would like to express my most sincere appreciation to my thesis advisor, Dr. Yizhi Jane Tao. This endeavor would not have been possible without her kindness and generous support. I am incredibly grateful for her patience and advice in my experimental designs and refinements. Her scientific expertise helps to set up many quality controls, essential for this first expression system of human picobirnavirus.

I could not have undertaken this journey without the help and suggestions from my thesis committee: Dr. Michael Gustin, Dr. George Bennett, Dr. Isaac Hilton, and Dr. Marcos de Moraes. Thank you all for your critical insights and encouragement for unbound exploration. Your comments stimulate plenty of creative thoughts in this work.

Many thanks to lab alums and present members in the Tao lab for being a lovely family at Rice. Special thanks to Dr. Matthew Ykema for sharing his experience in ultracentrifugation, TEM, and Nanosight equipment. I greatly appreciate Dr. Ying Zhou for helping me troubleshoot RT-qPCR experiments. I would like to acknowledge lab alumnus Dr. Junhua Pan for collecting electron cryo-microscopy images in Boston during the COVID pandemic, although we have not met in person.

I would like to extend my sincere thanks to Dr. Minna Poranen and Dr. Gabija Ziedaite-Ojala for collaboration and advice, Dr. Matthew Meyer for TEM training, Dr. Miguel Betancourt and Dr. Bo Chen for ultracentrifuge training, Dr. Ian Campbell for CD spectrometry training, and Dr. Li Li for performing all protein identity MS sequencing experiments.

My warmest thanks would go to my family: to my parents, Bo Fu and Qiong Song, to my grandparents, to my uncles and aunts, and cousins. Though most live thousands of miles away in China, their prayers continuously strengthen me. My love and gratitude would also go to my dear husband, Jingqi Pei. Thank you for supporting, encouraging, and believing in me.

Lastly, I would like to mention all the great opportunities at Rice. Thanks to IdeaLaunch and Lilie for entrepreneurship training, Rice office of technology transfer

(OTT) for patent and licensing training, and many people I interviewed for sharing their problems and solutions about nanoparticles.

Contents

List of Figures

List of Movies

List of Tables

Nomenclature

ICTV	International Committee for the Taxonomy of Viruses
CPV	Cypovirus
RRSV	Rice ragged stunt virus
RLR	RIG-I-like receptors
TLR3	Toll-like receptor
PKR	Protein kinase R
OAS	Oligoadenylate synthase
RNase L	Ribonuclease L
BLRP1	NOD-, LRR-, and pyrin domain containing 1
GSDMD	Gasdermin D
mRNA	Messenger RNA
dsRNA	Double-stranded RNA
+ssRNA	Positive-sense single-stranded RNA
-ssRNA	Negative-sense single-standed RNA
RBS	Ribosomal binding sequence
TEC	Transcriptional enzyme complex
CryoEM	Electron cryo-microscopy
PS	Packaging signal
hPBV	Human Picobirnavirus
rPBV	Rabbit Picobirnavirus

chPBV	Chicken Picobirnavirus
ORF	Open Reading Frame
CP	Capsid Protein
RdRP	RNA-dependent RNA polymerase
hPBV1	Viral RNA segment 1 of hPBV
hPBV2	Viral RNA segment 1 of hPBV
GVHD	Graft-versus-host disease
TID	Type-I diabetes
PCR	Polymerase chain reaction
EtBr	ethidium bromide
AmpR	Ampicillin resistance
KmR	Kanamycin resistance
SpecR	Spectinomycin resistance
IPTG	Isopropyl β-D-1-thiogalactopyranoside
VLP	Virus-like particle
UTR	Untranslated region
RT	Reverse transcription
TEM	Transmission electron microscopy
CD	Circular dichroism
HDV	hepatitis delta virus
WT	Wild type
IP	Immunoprecipitation

CcFV-1	*Colletotrichum camelliae* filamentous virus 1
MTase	Methyltransferase

Chapter 1

Introduction

1.1. Double-stranded (ds)RNA virus

As the name indicates, the dsRNA viruses comprise segmented dsRNA as their genomes. Though dsRNA viruses can infect many hosts, *Cystoviridae* is currently the only known prokaryotic dsRNA virus family (Table 1.1)[1–10]. A remarkable feature of all dsRNA viruses is their ability to transcribe and replicate viral RNA genome via a packaged RNA-dependent RNA polymerase (RdRP) inside the capsids. The transcriptionally self-sufficient characteristic is beneficial to surpass the host anti-viral mechanism activated by dsRNA presence[11]. In general, dsRNA viruses sequester their genomes inside capsids throughout their infection and replication life cycles. Therefore, dsRNA viruses share some structural and biochemical properties, especially a standard viral replication strategy[12,13]. Most dsRNA viruses possess non-enveloped icosahedral capsids. Their capsids comprise one to three protein layers and enclose up to 12 segments of the RNA genome. Each packaged RdRP is organized near capsid pores at vertices of symmetry axes to perform efficient transcription and replication. Thus, most dsRNA viruses are structurally similar in the innermost capsid layers and capsid-associated enzymes but diversify in their outer capsid layers and corresponding virion organizations (e.g., the triangulation number T). For example, under the *Reovirales* order (i.e.,

reoviruses), the *Sedoreoviridae* family has almost spherical and smooth surfaces due to lacking large surface projections on the subviral particle, while the *Spinareoviridae* family has spikes or turrets at five-fold vertices. Among turreted reoviruses, CPV has the simplest capsid structure, which lacks an outer capsid layer composed of trimers, while the *orthoreovirus* core is surrounded by 200 trimers of membrane penetration protein μ1[14,15]. In contrast, RRSV has only five peripheral trimers around each turret seated at the Q trimer position, like an architectural intermediate from CPV to *orthoreovirus*[16]. Hence, it is proposed that the turreted and non-turreted reoviruses share a common ancestor and separate further potentially via adaptation to their respective hosts.

Family	Number of genome segments	Type of virus particle	Host
Birnaviridae	2 (co-packaged)	~ 60 nm in diameter icosahedral, single shell, non-enveloped	Vertebrates (excluding mammals), invertebrates
Cystoviridae	3 (co-packaged, equimolar)	~ 85 nm in diameter, typically two concentric, icosahedral protein layers	Gram-negative Bacteria (*Psudomonas*)
Chrysoviridae	3-7 (separately packaged)	~ 40 nm in diameter, non-enveloped	Fungi, plants, and possibly insects
Megabirnaviridae	2 (likely packaged individually)	~ 52 nm in diameter, non-enveloped	Fungi
Paritiviridae	2 (separately packaged)	~ 25-43 nm in diameter, icosahedral protein capsid	Fungi, plants, protozoa
Picobirnaviridae	**2 (co-packaged)**	**~33-37 nm in diameter, non-evenloped, spherical virion**	**Vertebrates and invertebrates**
Polymycoviridae	4 (up to 8)	Non-conventionally encapsidated dsRNA, coated with viral protein	Fungi (ascomycetes and basidiomycetes), oomycetes
Quadriviridae	4 (separattely packaged)	~ 45 nm in diameter, non-enveloped	Fungi
Sedoreoviridae	10-12 (co-packaged)	~ 60-100 nm in diameter, non-enveloped, 1-3 concentric capsid layers	Mammals, birds, crusttaceans, arthropods, algae, and plants
Spinareoviridae	9-12 (co-packaged)	~ 60-85 nm in diameter, non-enveloped, 1-3 concentric capsid layers	Mammals, aquatic animals, birds, reptiles, arthropods, fungi, and plants

Table 1.1: Ten distinct families of dsRNA viruses. The table is summarized from ICTV reports by 2022. *Picobirnaviridae* is highlighted in bold.

1.1.1. General Life Cycle of dsRNA viruses

The presence of dsRNA is usually associated with viral infections either from the viral genome (i.e., dsRNA viruses) or an intermediate product during the viral genome replication (i.e., (+)RNA viruses)[12]. Accumulation of dsRNA can also appear from replication or transcription by-products of (-)RNA and dsDNA viruses[11]. Therefore, host antiviral strategies frequently recognize dsRNA by dsRNA-binding sensor proteins, such as RIG-I-like receptors (RLRs) in the cytoplasm and Toll-like receptor 3 (TLR3) in endosomes[17,18]. Upon detecting dsRNA presence, their oligomers activate respective

downstream adaptors of signaling pathways and eventually stimulate the host to produce type I interferons and other proinflammatory cytokines. Protein kinase R (PKR) dimerizes after recognizing dsRNA, auto-phosphorylates, and then phosphorylates eIF2α, a key translation initiation factor, eventually inhibiting global protein synthesis[19,20]. Oligoadenylate synthases (OASes) undergo conformational changes when binding dsRNA, synthesize 2'-5'-linked oligoadenylate as a soluble messenger to activate ribonuclease L (RNase L), which degrades cytosolic host and viral RNAs and finally suppresses protein synthesis[21,22]. NOD-, LRR- and pyrin domain-containing 1 (NLRP1) recognizes dsRNA by RNA duplex structure and length, then releases C-terminal UPA-CARD fragments, which forms inflammasomes and activates caspase 1[23]. Activated caspase 1 cleaves precursors of inflammatory cytokines, e.g., IL-1β[23,24]. Alternatively, active caspase 1 can also cleave pore-forming protein gasdermin D (GSDMD), which forms pores in the plasma membrane and activates the lytic cell death[23]. Even in prokaryotic cells, immune sensors are essential for host defense to distinguish between self and foreign nucleic acids. Like the RNase III family in eukaryotic cells (e.g., human Dicer), bacteria dsRNA-specific RNase III recognizes long dsRNA with little sequence specificity and cleaves them at preferred sites[25,26]. In addition, a genome mining analysis in the microbial pangenome showed novel bacterial defense systems encoded proteins comprising helicases, nucleases, and nucleic acid binding domains[27].

To avoid triggering the host antiviral mechanism, dsRNA viruses seal their dsRNA genome inside intact viral capsids throughout their replication cycle (Fig. 1.1). After viral infection, viral messenger RNAs (mRNA), which are the positive-sense single-stranded RNA (+ssRNA), are transcribed by the virion-associated RdRPs and then release into the cytoplasm through the exit channels on the capsid[28,29]. The released viral mRNAs translate viral proteins using the host protein synthesis system. In other words, mRNAs of a prokaryotic dsRNA virus recruit bacterial ribosomes by inherent ribosomal binding sequences (RBSs, a.k.a. Shine-Dalgarno sequences). In particular, a -1 frameshift in bacteriophage φ6, which generates a termination-initiation overlapping codon (UAAUG), enables protein translation of the genes without an upstream RBS[30]. By contrast, eukaryotic dsRNA viruses cap the 5' termini of each viral mRNA transcript with a methylated guanosine cap. For example, *Spinareoviridae* houses capping enzymes in

turrets on the twelve five-fold vertices of the innermost capsid layer. Their packaged viral RdRPs, which usually are coupled with NTPases as transcriptional enzyme complexes (TECs), are anchored by capsid shell proteins at the interior of these five-fold vertices[14,31–33]. Contrastingly, non-turreted reoviruses, e.g., rotaviruses, encapsulate their capping enzymes, which are enzymatically active *in vitro* and are proposed to stay near the transcript exit tunnel of packaged RdRPs[29,34,35]. After all viral proteins are translated, the packaging signal (PS) sequences on viral mRNAs, usually found in the untranslated regions, are their tickets to be specifically encapsulated into nascent capsids during viral assenbly[36]. Viral mRNAs serve as viral genome templates to allow RdRPs to synthesize the negative-sense single-stranded (-ss) RNA inside the newly assembled particles, converting the viral genome to its dsRNA form. The mature progeny viruses are finally released from the host and ready for the next infection cycle.

Since PBVs cannot yet be propagated in the laboratory, the life cycle of hPBV is proposed based on the general life cycle of a dsRNA virus (Fig. 1.1). Previous studies showed that hPBV capsid protein (CP) was arranged in a T=1 icosahedral capsid[37,38]. In
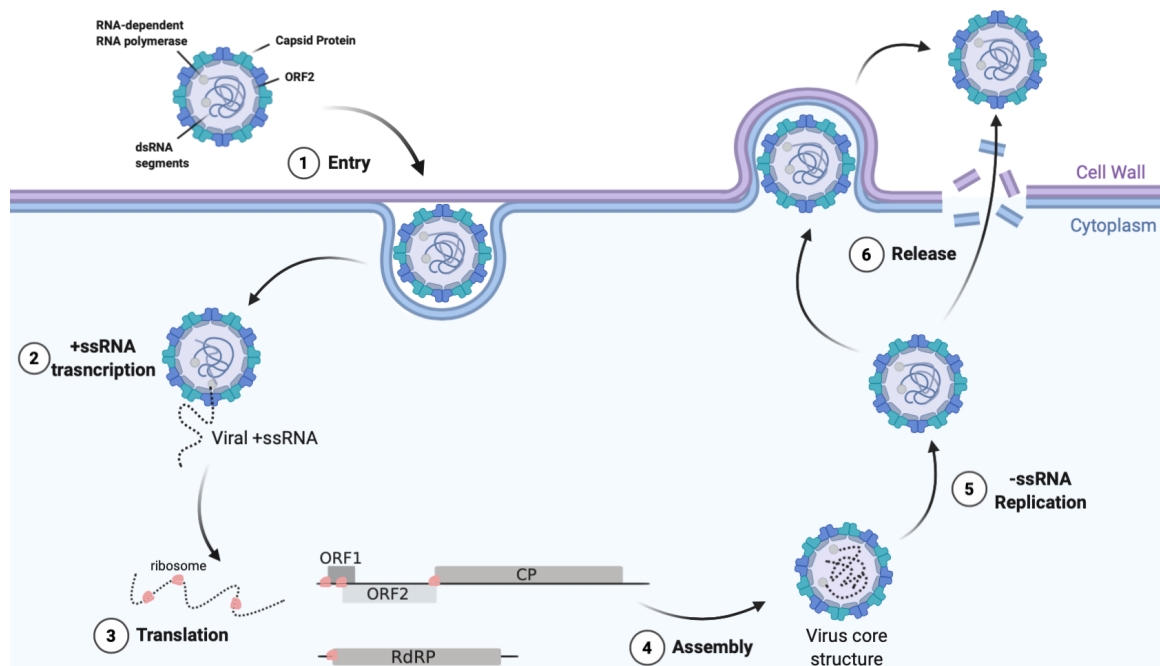
Figure 1.1.1: General life cycle of a typical dsRNA virus. A proposed dsRNA virus life cycle is demonstrated with a prokaryotic host cell. The viral genomes shown in the figure are the two RNA segments of the hPBV genome. The proposed hPBV viral structure represents a general dsRNA virus. Created with BioRender.com.

the proposed hPBV viral structure, 120 copies of CP form the single-layered hPBV capsid, encapsulating both dsRNA genome segments and RdRP(s) (Fig. 1.1).

1.1.2. RdRP transcription and replication mechanism of dsRNA viruses

To protect the dsRNA genome in isolated viral capsids from host attack, dsRNA viruses are self-sufficient in transcription and replication, a remarkable feature compared with other RNA viruses. Therefore, RdRPs are the most conserved viral proteins among dsRNA viruses and share critical strategies in transcription and replication. To clarify, "replication" refers to a process in which RdRP uses +ssRNA as a template and synthesizes -ssRNA. In contrast, "transcription" refers to a process in which RdRP uses dsRNA as a template and produces +ssRNA. The capability of dsRNA RdRPs to perform transcription and replication inside the viral capsids implies that both ssRNA and dsRNA can be taken as a template, making a dsRNA or ssRNA product, respectively.

Based on the *in situ* cryoEM structures of several reoviruses and bacteriophage φ6, RNA genomes are extended as rod-like densities featured with minor and major grooves and organized in multiple non-concentric layers with an approximate icosahedral D3 symmetry[14,29,31–33,39]. The termini of RNA duplex are held by encapsulated RdRPs or TECs, which are prepared for active transcription upon signals[31–33]. RdRPs are usually anchored by the N-terminal arms of capsid shell proteins at the twelve five-fold vertices[14,29,31,32,39,40]. In reoviruses, with up to twelve RNA segments, two five-fold vertices at the southern tropic remain empty in CPV, one unoccupied northern vertex in ARV, one random unfilled vertex in rotavirus, and two vacant vertices with each on either side in MRV[14,29,31–33]. The copy number of encapsulated RdRPs in bacteriophage φ6 viruses varies, which depends on the available RdRP concentration during viral assembly[41,42]. It is believed that φ6 RdRPs migrate from three-fold vertices in procapsids to five-fold vertices in mature capsids under an unknown mechanism, as RdRPs interact with capsids via hydrophobic interactions and salt bridges[39,41,43–45].

A representative transcription mechanism of dsRNA viruses can be summarized into three steps. Turreted reovirus CPV is used here as an example. The structure of CPV RdRP can be divided into an N-terminal domain, a core polymerase domain, and a C-terminal bracelet domain[31]. In the quiescent state, the RNA cap of terminal genomic RNA is held by the N-terminal domain of RdRP[31,33]. After ATP and SAM bind to the turret proteins (TP), CPV capsids expand at the five-fold vertices[40]. Conformational changes of capsid proteins result in the refolding of the RdRP bracelet domain, which uncovers the template entry tunnel and allows dsRNA duplex separation[40]. In the second step, the capsid-proximal module of the RdRP bracelet domain unveils the transcript exit tunnel, allowing elongation initiation[40]. Lastly, the thumb subdomain of the polymerase core opens and relaxes the bracelet domain for continuous elongation[40]. Template and non-template RNA strands re-anneal outside of the template exit tunnel, and the nascent transcript releases from the transcript exit tunnel towards capsids[40]. If transcription
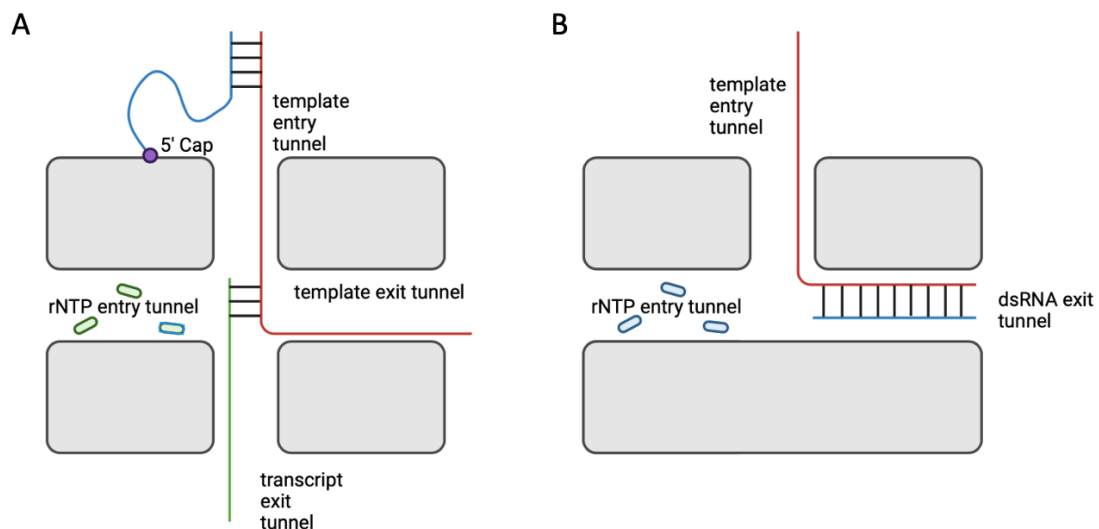
Figure 1.1.2: Simplified RdRP structures for transcription and replication. (A) RdRP in the transcribing state. RdRP N-terminal domain separates the dsRNA duplex like a helicase and holds the 5' cap (i.e., the non-template +ssRNA). The -ssRNA template enters through the template entry tunnel. The nascent transcript is synthesized in the polymerase core and released through the transcript exit tunnel. Template RNA leaves through the template exit tunnel and re-anneals with the non-template strand outside of RdRP, guided by positively charged residues on RdRP outside surface. (B) RdRP in the replicating state. The +ssRNA template comes in through the template entry tunnel, and the dsRNA product leaves through the dsRNA exit tunnel facing towards the center of viral particles. Created with BioRender.com.

aborts, the transcript exit tunnel remains open to release the interrupted transcript, avoiding its accumulation inside capsids[40].

By contrast, in the replication mechanism, the +ssRNA template enters the RdRP template entry tunnel, and the dsRNA product emerges from the dsRNA exit channel (Fig. 1.1.2). However, no *in situ* replication structure of dsRNA viruses is available. The procapsid structure of bacteriophage φ6 shows that RdRPs at three-fold vertices are orientated with the substrate tunnel pointing outwards, template entry tunnel tending towards sideways, and the product exit tunnel covered by RdRP C-terminal domain facing towards the procapsid center, which is presumably at the quiescent state prepared for replication[44]. The dsRNA exit tunnel is likely wider to allow dsRNA pass-through, compared with the template and transcript exit channels in the transcription mechanism, which accommodate ssRNA. Based on the proposed replication mechanism, the transcript exit tunnel facing toward capsids in the transcription mechanism should be fully covered.

1.2. Human *Picobirnavirus* (hPBV)

1.2.1. Discovery History and Epidemiology of PBV

Picobirnavirus (PBV) is the only genus under the family *Picobirnaviridae*. The name picobirnavirus is derived from its size and genome characteristics, where "pico" refers to the small size of the virus particle (i.e., ~ 35 nm in diameter) and "birna" describes the two-segmented dsRNA genome. Human PBV (hPBV) is the type species of the *Picobirnavirus* genus.

PBV was first discovered during acute gastroenteritis outbreaks in free-living rats (*Oryzomys nigripes*) from Brazil in 1988[46]. From 1988 to date, PBV has been detected in the stool samples of various animal species on each continent (Fig. 1.2.1.1)[47,48]. An early study analyzing the dsRNA viruses in pig feces by RNA PAGE showed that PBVs were more frequently detected in diarrheic pigs than non-diarrheic pigs (15.3% and 9.6%, respectively)[49]. Since then, hPBV is often associated with "gastroenteritis" and "diarrhea" in children and immunocompromised patients. Later, in the screenings of human diarrhea

samples, hPBV was also detected in those samples that were negative for known enteric causes[42,49,50]. However, according to the accumulated detection results in the feces of various animals in different areas, there is no conclusive or direct correlation between PBV and diarrhea[47].
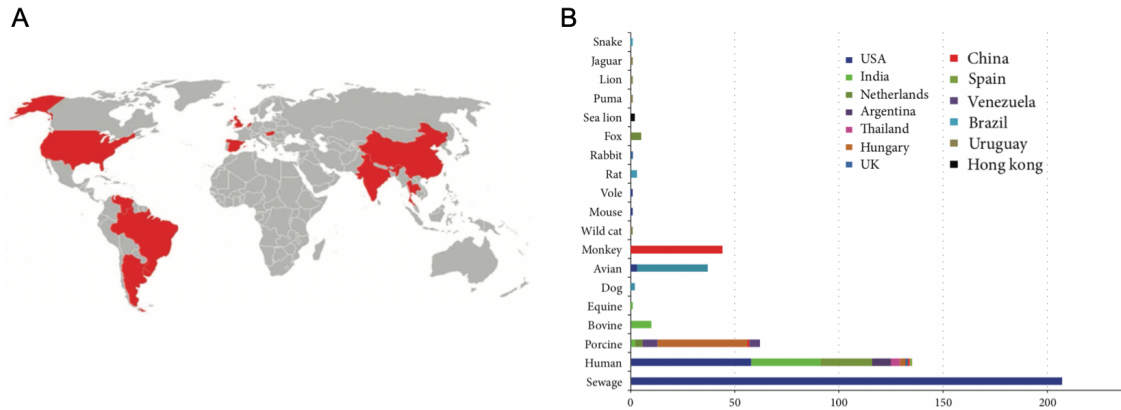
Figure 1.2.1.1: Geographic and Species-wide distribution of PBV. (A) Global distribution of PBV, the red color presents countries where PBV has been detected in any species, including sewage. (B) Global species-wise distribution of PBV based on the sequence deposited in the NCBI database. (Adapted from Malik et al. 2014)

In addition, PBV is proposed to play a synergetic role in the viral infection associated with primary enteric causes. A positive correlation was detected between PBV and other enteric pathogens (e.g., rotaviruses, astrovirus, *Salmonella*) in diarrhea animals[51–53]. Furthermore, the presence of PBV in HIV patients is significantly associated with gastroenteritis[54,55]. Interestingly, two major PBV genogroups are associated with HIV infection statuses[56]. PBVs are generally categorized into five genogroups (GI-GV) based on the RdRP sequence similarity (Fig. 1.2.1.2). PBVs detected in human patients are mainly classified into GI and GII genogroups[53,54,57]. The reference strain of GI is hPBV strain 1-CHN-97 from a non-HIV-infected gastroenteritis patient, while GII refers to hPBV strain 4-GA-91 from an HIV-infected gastroenteritis patient[56,58]. A genetic hPBV variant in the feces of a diarrhea patient reported in 2014 shares only 19.4-26.1% amino acid similarity with GI and GII PBVs, classified as GIII[59]. GIV and GV clusters are reported from a virome analysis in diseased horses[60]. In general, the amino acid identities of PBVs range from 44.8% to 97.1% within the same genogroup and 21.6% to 30.8% between different genogroups[60]. By 2014, 83.11% of 515 identified PBV sequences were classified as G-I and 2.52% as G-II[48]. A recent study suggested that

hPBV abundance is correlated to the occurrence of severe enteric graft-versus-host disease (GVHD) in the early post-transplantation of hematopoietic stem cells[61]. Another recent report on the gut virome in pregnant women found that PBVs are more likely to be detected in women with type I diabetes (TID) than those without TID[62]. It is hypothesized that the PBV prevalence may be a biomarker of immunosuppression.

To sum up, PBV has been widely detected in many geographical regions and a broad range of host species, including humans, mammals, reptiles, marine animals, birds, algae, and even in sewage (Table 1.1 and Fig. 1.2.1). However, the symptom of PBV infection and the pathogenesis of PBV remain unknown. No successful PBV propagation has been reported in any cell culture or animal model. Hence, a cellular model of PBV infection is urgently needed for further research.
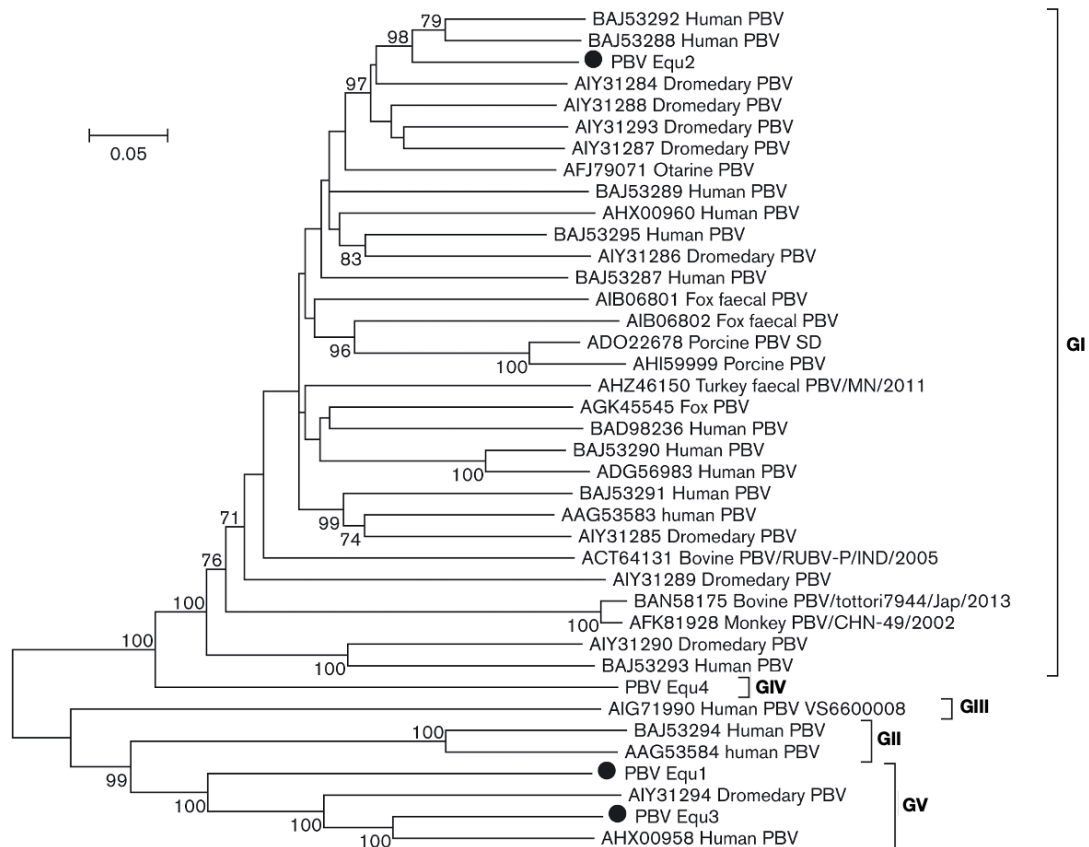
Figure 1.2.1.2: An unrooted phylogenetic tree of PBV GI-GV. The analysis is based on 450 amino acids of the RdRP protein sequence. PBV strains isolated from the same host species can distance far from each other in the phylogenetic tree. PBV Equ1-4 are isolated from horse feces. (Adapted from Li et al. 2015).

1.2.2. Host Theory of PBV

1.2.2.1. Eukaryotic host theory

Since PBV was isolated from the feces of different animals, like rotavirus and astrovirus, it was first proposed to be a eukaryotic virus. The nomenclature of PBV species always indicates its discovery origin. For example, rabbit picobirnavirus (rPBV) was initially isolated from rabbits, chicken picobirnavirus (chPBV) from chickens, and human picobirnavirus (hPBV) from humans.

Early studies of PBV genomic sequences from different animals suggested effective interspecies transmission of PBV[57,63–67]. Porcine PBV strains are diverse in genome profile and highly related to GI human PBVs, especially to a Hungarian human PBV strain[63,64]. Similar studies in GI porcine PBVs showed that porcine PBVs in Hungary, Venezuela, and Argentina are closely related to human GI PBV, indicating viral circulation in humans and pigs, suggesting likely zoonotic potential[57,64,65]. Another similar study later showed that horse PBV strains are related to an Indian human PBV strain and a few Hungarian porcine PBV strains[66]. These reports explain the wide range of PBV hosts by efficient host-host transmission[67].

Moreover, based on the synergetic role of PBV in viral gastroenteritis, it is proposed that PBV is a eukaryotic virus, which explains its presence with other eukaryotic pathogenetic viruses. Besides the co-existence of PBV with other enteric viruses, HIV, and GVHD mentioned earlier, a recent study in hospitalized patients with acute respiratory illness during the COVID pandemic reports various respiratory-tropic pathogens in PBV-positive patients[68]. However, among these respiratory-tropic pathogens, no specific shared respiratory pathogenetic bacteria as a potential co-infection host of PBV can be found among PBV-positive patients[68].

1.2.2.2. Prokaryotic host theory

Although PBV has been discovered in the feces of humans and many animals, another hypothesis suggests the possible nature of PBV as a bacteriophage. A recent analysis of 81 PBV genome sequences revealed that the RBSs are abundantly enriched near the start of all predicted open reading frames (ORFs) (Fig. 1.2.2)[69]. Similar phenomena are common in prokaryotic viruses but rare in eukaryotic viruses, given that mRNA translation in prokaryotic cells relies on RBSs for 16S rRNA subunit recruitment. The functionality of the predicted RBSs was supported by the detection of C-terminal His-tagged peptides expressed under the RBS-like sequences from the chicken PBV (chPBV)[70]. Therefore, it is hypothesized that hPBV is a prokaryotic virus propagating in the human enteric tract.

Figure 1.2.2: RBS enrichment near the start of all predicted ORFs. (A) 5' UTR analysis of known and novel PBV. RBS sequences are highly enriched upstream of predicted ORFs (Adapted from Krishnamurthy and Wang, 2018). (B) 5'UTR RBS predictions of hPBV (Hy005102) RNA segment in this study. RBS sequences are highlighted in pink, and the start codons are highlighted in purple.

In addition, the assumption of PBV as a prokaryotic virus also explains its wide distribution in different animals and the detection of PBV in respiratory tract samples, blood samples, and even in environments like sewage[48,60,71]. Besides, the bacteriophage nature also demonstrates the detection of multiple distinctive PBV strains within individuals[65,68,72]. For instance, a report in 2011 detected a mixed infection of four PBV strains from both GI and GII in a diarrhea child from India[73].

1.2.2.3. Mitochondrial host theory

However, the above does not exclude the possibility that some PBV strains may be eukaryotic viruses infecting mitochondria. Yinda et al. identified classical and novel PBV RdRP sequences in human and fruit bat fecal samples[74,75]. Those novel PBV-like RdRP sequences were translated using invertebrate mitochondrial genetic codes, known in mitoviruses[74,75]. The alternative codon usage of invertebrate mitochondria differs from the universal genetic codes in various ways, including reassigning the TGA stop codon to another amino acid, such as tryptophan (TGG)[74,76,77]. Based on the codon usage bias of different known mitochondrial genome sequences, the novel PBV-like RdRP sequences were speculated to have a life cycle similar to mitoviruses, dsRNA viruses infecting fungal mitochondria[76–78]. The genome of mitoviruses only encodes an RdRP but does not form encapsulated viral particles. Later studies detected more novel PBV-like sequences using alternative mitochondrial codons in mongooses and wild birds[78,79]. Notably, an unsegmented novel PBV-like virus from birds was translated using a yeast mitochondrial genetic code[79].

1.2.3. Structural characterization of hPBV proteins in previous studies

PBVs are recognized mainly by their bi-segmented dsRNA genome, especially on RNA PAGE by silver staining[46]. In general, PBV has a 2.3-2.6 kb segment 1 and 1.5-1.9 kb segment 2[80]. Among PBV strains, the viral CP gene locates in segment 1, and segment 2 only encodes the RdRP (Fig. 1.2.3.1). In most PBV segment 1 RNA sequences, one or two functionally uncharacterized ORF(s) are located upstream of the CP-coding

sequence. Recently, unsegmented PBV sequences have also been identified in wild animals[79,81].

The hPBV strain characterized in this work is the Hy005102 strain, a GI hPBV strain isolated from the stool of an infant with acute non-bacterial gastroenteritis in Thailand[82]. This strain is the only hPBV with a full-length cDNA sequence deposited in the NCBI database. In addition to the 552-amino-acid CP, viral RNA segment 1 (hPBV1) is also predicted to encode a 39-amino-acid ORF1 and a 224-amino-acid ORF2 (Fig. 1.2.3.1). Both ORF1 and ORF2 are predicted proteins with unknown functions. ORF2 and CP sequences have a -1 frameshift, but each has an upstream inherent RBS sequence (Fig. 1.2.2 B). A 24.9 kDa protein was once observed in a previous report when expressing the full-length hPBV cDNA in the cell-free system[82].

Previous studies resolved the CP structure of rabbit PBV (rPBV) by crystallography and that of hPBV by electron cryo-microscopy (CryoEM)[37,38]. CP of rPBV exhibits a self-cleavage between residue 65 and 66 at the N-terminus, mapped to residue 45 and 46 in hPBV (Fig. 1.2.3.2 A)[37,38]. In rPBV, the self-cleaved CPs automatically dimerize through the N-terminal interactions[38]. The intricate dimer interface was mediated by the rest of the N-terminal 55 residues (i.e., amino acids 66-120). Two CP dimers form a diamond tile, and thirty tiles assemble into the rPBV capsid (Fig. 1.2.3.2 B). However, the hPBV CP did not exhibit any self-cleavage activity in the *E. coli* expression system[37,83]. The uncleaved hPBV CPs readily dimerize and assemble into capsids (Fig. 1.2.3.2 C), which raises questions about the functional relevance of the self-cleavage activity observed in rPBV CP. Due to its positively charged N-terminus, the assembled CP can randomly package rRNA and ribosomal proteins[37]. At the 5-fold axes, where RdRP frequently seats in most dsRNA viruses[29,31,32], a 9-Å pore is formed by the
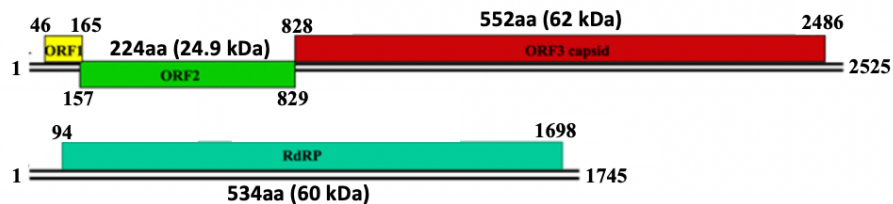
Figure 1.2.3.1: The genome organizations of hPBV. The genome of the hPBV Hy005102 strain is shown as a representative PBV genome. The viral RNA segment 1 is 2.5 kb and encodes ORF1, ORF2, and capsid protein, while segment 2 encodes RdRP.

N-termini of the surrounding CPs and a simple conformational change of CP N-terminal loop (131-138 residues) results in a 23-Å pore opening (Fig. 1.2.3.2 D)[37]. It implies that hPBV may also organize its encapsulated RdRPs at five-fold vertices like other dsRNA viruses. Hence, the capsid pore opening via the conformational change allows RNA transcript to release in the transcribing state.

The RdRP of hPBV has been characterized by X-ray crystallography, and a 2.4 Å resolution structure is publicly available. *In vitro* polymerase assays indicated that the isolated hPBV RdRP performs a *de novo* semi-conservative replication/transcription mechanism[83]. While the core polymerase domains (i.e., thumb, palm, and fingers) are well conserved, hPBV RdRP has a highly flexible loop structure in its active site, which is responsible for the *de novo* priming during the initiation of RNA synthesis (Fig. 1.2.3.3). Moreover, hPBV RdRP was observed to preferably bind to its viral RNA conserved 5'-terminal viral RNA sequence, which has a low melting temperature that can
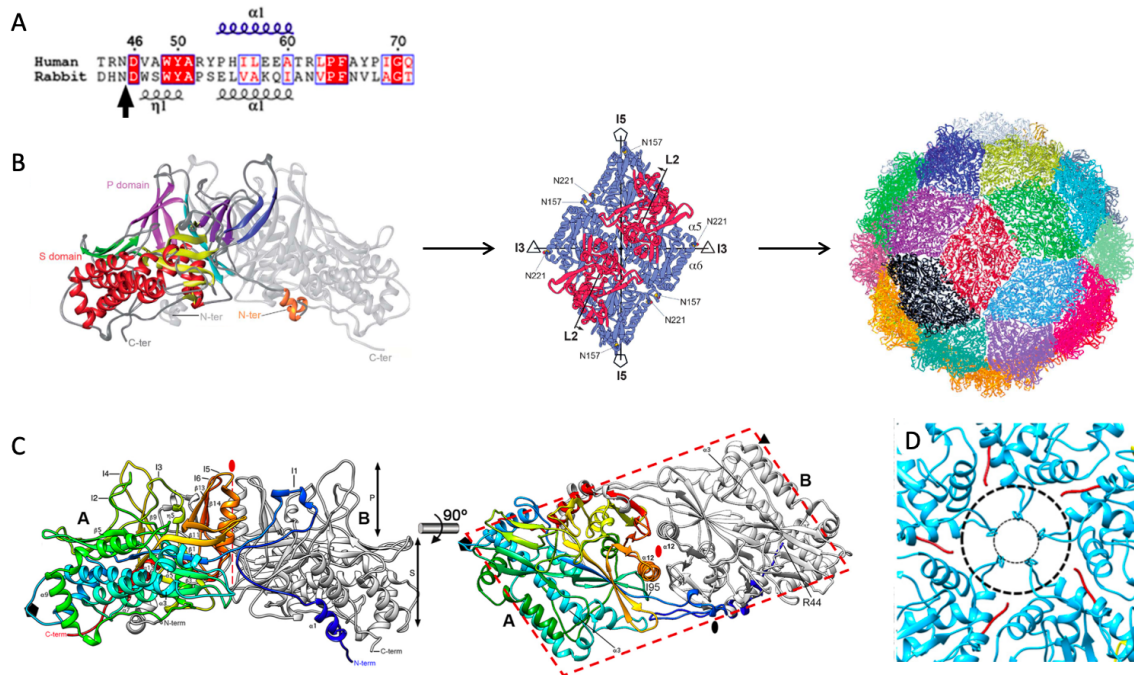
Figure 1.2.3.2: CP structure and capsid assembly in PBV. (A) Amino acid sequence alignment between rPBV and hPBV to demonstrate the predicted hPBV self-cleavage site based on the rPBV sequence. (B) rPBV CP structure and capsid assembly mechanism. CP self-assembles into dimers, two dimers form a diamond tile, and eventually thirty tiles assemble into the capsid. (C) The dimer structure of hPBV CP. (D) The 5-fold axes top view of hPBV capsids. A 9-Å pore is indicated by the inner dashed circle and a 23-Å pore by the outter dashed circle. (Ortega-Esteban et al. 2020 and Duquerroy et al. 2009)

facilitate transcription initiation. However, the previous experiments observed no direct interaction between the CP and RdRP. It is hypothesized that RdRP is packaged into the capsids via its interaction with the viral RNA termini.

1.2.4. Summaries and objectives
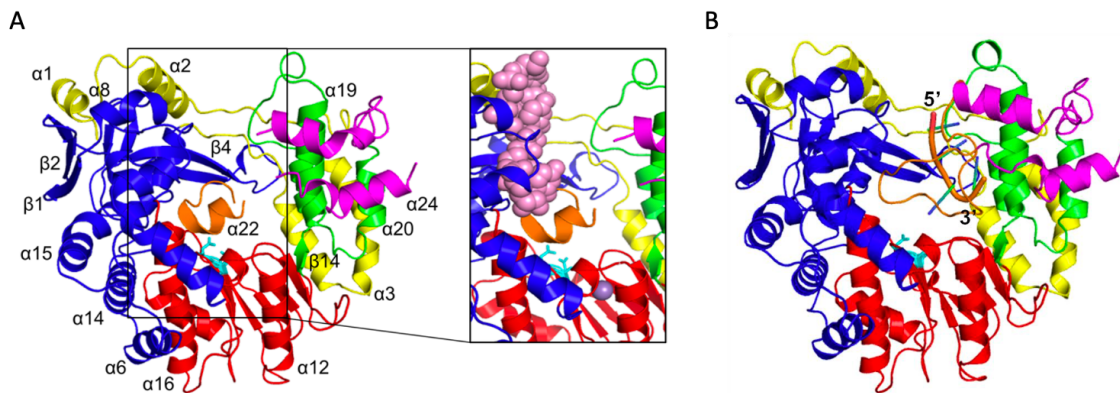
Figure 1.2.3.3: RdRP structure of hPBV. The conventional polymerase core is colored with the *thumb* in green, the *palm* in red, and the *fingers* in blue. The N-terminal domain of RdRP is colored in yellow, and the C-terminal domain is in magenta. The flexible C-terminal loop is highlighted in orange. (A) Structure of hPBV RdRP, a zoomed view with the superimposed oligonucleotide from φ6 RdRP (PDB 1HI0). (B) Structure of hPBV RdRP with single-stranded DNA molecules. Both structures demonstrate the estimated RNA termini binding position and its interaction with the C-terminal loop. (Adapted from Collier et al. 2016)

As a member of dsRNA viruses, picobirnaviruses (PBVs) have been detected in various hosts, including humans and different animal species. However, pathogenicity and pathogenesis associated with PBV infection remain unknown due to the lack of cell or animal models for PBV propagation in the laboratory. The uncertainty in the host identity of PBVs presents a significant obstacle in developing a PBV culture system in the laboratory. Based on the analysis of PBV genomes, where RBSs are enriched in the upstream region of all predicted ORFs, and the fact that PBVs isolated from the same host species or even the same host could be phylogenetically far related, I hypothesized that PBVs are prokaryotic viruses, i.e., bacteriophages.

This work aimed to establish a plasmid-based expression system for recombinant human picobirnavirus (hPBV) to test the function of predicted RBSs and investigate its viral assembly mechanism. The +ssRNA of each dsRNA segment was expressed from

plasmids in an expression host, *E. coli* Rosetta strain. *E. coli* Rosetta 2 (DE3) cells contain a plasmid to encode eukaryotic tRNAs for rare codon translation in *E. coli*, which facilitates overcoming potential codon bias of the expressed hPBV RNA segments. As expected, all four predicted ORFs were translated, suggesting that their upstream RBSs are functional and that hPBV is a prokaryotic virus. Purified recombinant hPBVs from the expression system were characterized by their RNA and protein components. The identity of recombinant hPBV was confirmed by its morphology under transmission microscopy and protein sequencing by mass spectrometry. I then investigated the incorporation mechanism of the encapsulated viral RNA and proteins. At last, co-immunoprecipitation assays of recombinant hPBVs mixed with *E. coli* lysate indicated that *E. coli* is likely not a native host of hPBV. A similar approach may be used to search for the natural host(s) of PBVs in future studies.

Chapter 2

Methods

The principle of our recombinant hPBV expression system is expressing the exact viral RNA segments to mimic the viral infection using a heterogeneous cell line. To ensure the expression of viral RNA segments with precise termini, the cDNA of each hPBV RNA segment was flanked by a T7 promoter and HDV ribozyme sequence at the 5'- and 3'-ends, respectively. Given that the two hPBV viral RNA segments encode at most four predicted viral proteins, plasmids were constructed for viral RNA(s) expression, viral protein(s) expression, or their combinations. Each plasmid was first individually expressed in *E. coli* Rosetta^{TM} 2 (DE3) cells for expression characterization. To study the assembly mechanism of hPBV, recombinant hPBV was purified and characterized from the combinations of different viral RNA and protein co-expression. The identity of recombinant hPBV was confirmed by its protein profile from MS sequencing and its morphology under transmission electron microscopy (TEM). The encapsulated RNA molecules from immunoprecipitation assays were characterized by RNA extraction and strand-specific RT-qPCR. Co-immunoprecipitation assays and Western blots from different expression combinations demonstrated the incorporation mechanism of packaged viral proteins. Purified recombinant hPBV was also visualized under Nanosight to validate its size distribution and RdRP incorporation via a GFP tag. This chapter summarizes all lab techniques and methods used in this work.

2.1. Plasmid construction

The cDNA of hPBV RNA segment 1 (AB186897) was synthesized in a pUC57 plasmid from GeneScript. The cDNA of hPBV RNA segment 2 (AB186898), obtained from David Wang's lab, was flanked by a T7 promoter and a hepatitis D virus (HDV) ribozyme sequence at 5'- and 3'-ends in the pEXT20 vector. Both cDNA sequences of the hPBV genome were confirmed by the Sanger sequencing from Genewiz. Dr. Aaron Collier provided the plasmids and sequences for codon-optimized CP and RdRP from his previous work at Rice Univeristy[83]. Empty vectors of pET28a, pETDuet, pET19b, pBAD18, and pEXT20 backbones were all available in the lab, obtained from ATCC. The pET-SpecR plasmid backbone was constructed from pET28a by swapping the kanamycin antibiotic resistance gene to a spectinomycin antibiotic resistance gene from the pJEC103 vector. The original pJEC103 plasmid was obtained from Dr. James Chappell's lab. Similarly, pBAD18-KmR substituted the original ampicillin resistance gene with a kanamycin resistance gene from the pET28a vector.

DNA fragments of inserts and vector backbones were all amplified by polymerase chain reactions (PCR) with Q5® high-fidelity DNA polymerase (NEB M0491L). Primers were obtained from Integrated DNA Technologies (IDT). During PCR amplification, short fusion tags (e.g., a His or Strep tag) were inserted into the DNA fragments as part of the primer sequences in their overhang regions. The 5'-CATCATCACCATCACCAC was the sequence of a His tag and 5'-TGGTCGCATCCGCAGTTCGAGAAG for a Strep tag. Each 50 μl Q5 PCR reaction comprised: 1 μl of the DNA template (10-20 ng/μl), 1.25 μl of a forward primer (25 μM), 1.25 μl of a reverse primer (25 μM), 1 μl of dNTP mix (10 mM), 10 μl of 5× Q5 reaction buffer, 10 μl of Q5 high GC enhancer, 25 μl of nuclease-free water, and 0.5 μl of Q5® high-fidelity DNA polymerase. Reagents were added into a PCR tube in the listed order. PCR reactions were carried out by a thermal cycler (Bio-Rad T100™) for 35 cycles of denaturation, primer annealing, and extension. The PCR cycling parameters for a regular PCR were 98 ℃ for 10 seconds, 55 ℃ for 30 seconds, and 72 ℃ for 30 seconds per kb. The PCR cycling parameters for a touch-down PCR to specifically amplify a short insertion fragment (< 2 kb) were 98 ℃ for 10 seconds, 65-55 ℃ for 30 seconds (decreasing 0.5 ℃ each cycle for 20 cycles), 72 ℃ for

30 seconds per kb, and then followed by regular PCR amplification of 35 cycles. All PCR products were confirmed by 2% agarose gel electrophoresis, stained with ethidium bromide (EtBr), and visualized under UV. The methylated DNA templates were removed by DpnI (NEB R0176L) digestion at 37 ℃ for 2 hours to avoid false positive results in the following transformation experiments. Salts and enzymes in DpnI-treated PCR reactions were removed by QIAquick PCR purification kit (QIAGEN 28104) or by QIAquick gel extraction kit (QIAGEN 28706) with gel slices from 1% agarose gel electrophoresis if other unspecific PCR products were observed in the previous 2% agarose gel electrophoresis. The purified DNA amplificants were eluted in nuclease-free water.

DNA fragments of inserts and backbone vectors were ligated by Gibson assembly (NEB Gibson Assembly® Mater Mix E2611L) or In-Fusion assembly (Takara In-Fusion® Snap Assembly mater mix 638948). Inserts and vectors were mixed in a molar ratio of 3:1 or 5:1 for longer (> 1kb) or short (< 1kb) inserts, respectively. Each 5 μl Gibson assembly cloning reaction comprised 2.5 μl of 2× Gibson Assembly® Mater Mix and 2.5 μl of DNA fragment mixture, including 12.5-25 ng of a vector. Each 5 μl In-Fusion assembly cloning reaction comprised 1 μl of 5× In-Fusion® Snap Assembly mater mix and 4 μl of DNA fragment mixture. Cloning reactions were incubated at 50 ℃ for 15 minutes, immediately chilled on ice for 3-5 minutes, and transformed into *E. coli* DH5α competent cells. The DH5α competent cells were transformed using a general chemical transformation protocol: 30-min ice incubation with DNA samples, 1-min heat shock at 42 ℃, 1-min chilling on ice, and 30- to 60-min recovery in LB or S.O.C. media at 37 ℃. The transformed DH5α cells were plated on selective LB agar plates and incubated at 37 ℃ for 24 hours. A single colony was picked and seeded to 5 ml of corresponding selective LB liquid culture, shaken overnight at 37 ℃. The cloned plasmids were eventually isolated by QIAprep Spin Miniprep Kit (QIAGEN 27106). Plasmid sequences were confirmed by Sanger sequencing (Genewiz). All plasmids used in this work are summarized in the following tables.

Name	Construct	Antibiotic Resistance	Note
pET19b-CP	Codon-optimized CP	Amp^R	From Dr. Aaron Collier
pET28b-CP	Codon-optimized CP	Km^R	From Dr. Aaron Collier
pET28b-GFP-CP	GFP fused codon optimized capsid protein	Km^R	From Dr. Aaron Collier
pET28a-ΔCP	Codon-optimized truncated Δ45-CP	Km^R	From Dr. Aaron Collier

Table 2.1.1: Plasmids for codon-optimized CP protein expression. Codon-optimized CP expression constructs were all under the RBS from the vector backbone as for regular recombinant protein expression. Amp^R: ampicillin resistance. KmR: kanamycin resistance.

Name	Construct	Antibiotic Resistance	Note
pET19b-GFP	Expression of GFP as controls	Amp^R	
pET28b-His-RdRP	Codon-optimized His-RdRP protein expression	Km^R	From Dr. Aaron Collier
pET19b-GFP-RdRP	Codon-optimized GFP-RdRP protein expression	Amp^R	
pET19b-RdRP-GFP	Codon-optimized RdRP-GFP protein expression	Amp^R	
pET19b-RdRP	Codon-optimized RdRP without any tag	Amp^R	
pBAD18-KmR-His-RdRP	Codon-optimized His-RdRP protein expression under arabinose induction	Km^R	

Table 2.1.2: Plasmids for codon-optimized RdRP protein expression. All protein expression constructs were under the RBS from the vector backbone. The GFP sequence was cloned from pET28b-GFP-CP in Table 2.1.1 from Dr. Aaron Collier.

Name	Construct	Antibiotic Resistance	Note
pET-SpecR-ORF2-FL	His-ORF2 protein expression	Spec^R	
pET-SpecR-ORF2-FL.1	His-ORF2(FL)-Strep protein expression	Spec^R	
pET-SpecR-ORF2(38-224)	ORF2 N-terminal truncates: His-ORF2 (38-224) protein expression	Spec^R	
pET-SpecR-ORF2(86-224)	ORF2 N-terminal truncates: His-ORF2 (86-224) protein expression	Spec^R	
pET-SpecR-ORF2(107-224)	ORF2 N-terminal truncates: His-ORF2 (107-224) protein expression	Spec^R	
pET-SpecR-ORF2(1-107)	ORF2 C-terminal truncates: His-ORF2 (1-107) protein expression	Spec^R	
pET-SpecR-ORF2(1-107).1	ORF2 N-terminal truncates: His-ORF2(1-107)-Strep protein expression	Spec^R	
pET-SpecR-ORF2(1-163)	ORF2 C-terminal truncates: His-ORF2 (1-163) protein expression	Spec^R	

Table 2.1.3: Plasmids for ORF2 protein expression. ORF2 was expressed under the RBS from the vector backbone. ORF2 sequence was directly cloned from hPBV1 without codon optimization. Spec^R: spectinomycin resistance.

Name	Construct	Antibiotic Resistance	Note
pETDuet-CP-ORF2(FL)	Protein co-expression of codon-optimized CP and His-ORF2 (FL)	Amp^R	
pETDuet-CP-ORF2(38-224)	Protein co-expression of codon-optimized CP and His-ORF2 (38-224)	Amp^R	
pETDuet-CP-ORF2(86-224)	Protein co-expression of codon-optimized CP and His-ORF2 (86-224)	Amp^R	
pETDuet-CP-ORF2(107-224)	Protein co-expression of codon-optimized CP and His-ORF2 (107-224)	Amp^R	
pETDuet-CP-ORF2(1-107)	Protein co-expression of codon-optimized CP and His-ORF2 (1-107)	Amp^R	
pETDuet-CP-ORF2(1-163)	Protein co-expression of codon-optimized CP and His-ORF2 (1-163)	Amp^R	
pETDuet-ΔCP-ORF2(FL)	Protein co-expression of codon-optimized Δ45-CP and His-ORF2 (FL)	Amp^R	
pETDuet-ΔCP-ORF2(1-107)	Protein co-expression of codon-optimized Δ45-CP and His-ORF2 (1-107)	Amp^R	

Table 2.1.4: Plasmids for CP and ORF2 protein co-expression. CP or Δ45-CP was cloned from the plasmids in Table 2.1.1 obtained from Dr. Aaron Collier. ORF2 was expressed under the RBS downstream from the second T7 promoter, while CP or Δ45-CP expression was under the first T7 promoter.

Name	Construct	Antibiotic Resistance	Note
pET28-ΔRBS-hPBV1	Wild-type hPBV segment 1 expression under T7 promoter	Km^R	
pET28-ΔRBS-hPBV1.1	hPBV segment 1 expression under T7 promoter. His-SUMO tagged ORF1.	Km^R	
pET28-ΔRBS-hPBV1.2	hPBV segment 1 expression under T7 promoter. His-tagged ORF2.	Km^R	
pET28-ΔRBS-hPBV1.3	hPBV segment 1 expression under T7 promoter. His-tagged CP.	Km^R	
pET28-ΔRBS-hPBV1.4	hPBV segment 1 expression under T7 promoter. Strep-tagged CP	Km^R	
pET28-ΔRBS-hPBV1.5	hPBV segment 1 expression under T7 promoter. ORF2 was removed by 1-107 deletion.	Km^R	
pETDuet-ΔRBS-hPBV1	Wild type hPBV segment 1 under the first T7 promoter in pETDuet vector	Amp^R	

Table 2.1.5: Plasmids for hPBV1 expression. The cDNA sequence of hPBV1 was flanked by the T7 promoter from the vector backbone at the 5'-end and an HDV ribozyme sequence at the 3'-end. The detailed sequences are in Appendix 2.

Name	Construct	Antibiotic Resistance	Note
pETDuet-hPBV1-ORF2(FL)	His-tagged ORF2 (FL) and hPBV1 co-expression	Amp^R	ΔRBS-hPBV1
pETDuet-hPBV1-ORF2(38-224)	His-tagged truncated ORF2 (38-224) and hPBV1 co-expression	Amp^R	ΔRBS-hPBV1
pETDuet-hPBV1-ORF2(86-224)	His-tagged truncated ORF2 (86-224) and hPBV1 co-expression	Amp^R	ΔRBS-hPBV1
pETDuet-hPBV1-ORF2(1-107)	His-tagged truncated ORF2 (1-107) and hPBV1 co-expression	Amp^R	ΔRBS-hPBV1
pETDuet-hPBV1-ORF2(1-163)	His-tagged truncated ORF2 (1-163) and hPBV1 co-expression	Amp^R	ΔRBS-hPBV1

Table 2.1.6: Plasmids for the co-expression of viral RNA segment hPBV1 expression with the full-length or truncated ORF2 protein. T7-hPBV1-HDV viral RNA expression cassette was cloned under the first T7 promoter of the pETDuet vector, where hPBV1 replaced the backbone RBS sequence for protein expression. ORF2 gene was located under the RBS downstream of the second T7 promoter of the pETDuet vector.

Name	Construct	Antibiotic Resistance	Note
pEXT20-T7-hPBV2	Wild-type hPBV segment 2 expression under T7 promoter.	Amp^R	From Dr. David Wang's lab.
pEXT20-T7-hPBV2.1	hPBV segment 2 expression, encoding His-RdRP	Amp^R	
pEXT20-T7-hPBV2.2	hPBV segment 2 expression, encoding GFP-RdRP	Amp^R	
pEXT20-T7-hPBV2.3	hPBV segment 2 expression, encoding RdRP-GFP	Amp^R	

Table 2.1.7: Plasmids for hPBV2 expression. An upstream T7 promoter was inserted with hPBV2, replacing the original pTac promoter in the pEXT20 backbone. Detailed sequences are in Appendix 2.

Name	Construct	Antibiotic Resistance	Note
pETDuet-ΔRBS-hPBV	Co-expression of wild type hPBV1 and hPBV2	Amp^R	
pETDuet-ΔRBS-hPBV.1	Co-expression of hPBV1 and hPBV2 encoding His-RdRP	Amp^R	
pETDuet-ΔRBS-hPBV.2	Co-expression of hPBV1 and hPBV2 encoding His-RdRP (GDD-GAA)	Amp^R	
pETDuet-ΔRBS-hPBV.3	Co-expression of hPBV1 encoding CP and hPBV2 encoding His-RdRP	Amp^R	
pETDuet-hPBV1-RdRP	Co-expression of hPBV1 and codon-optimized His-RdRP protein	Amp^R	ΔRBS-hPBV1

Table 2.1.8: Plasmids for the co-expression of hPBV1 and hPBV2 (or RdRP). T7-hPBV1-HDV cassette starts at the first T7 promoter, and T7-hPBV2-HDV locates at the second T7 promoter in the pETDuet vector. Codon-optimized His-RdRP was expressed under the vector backbone RBS sequence downstream from the second T7 promoter.

2.2. Viral RNA segment and protein expression

Expression plasmids or combinations of plasmids were (co-)transformed into *E. coli* Rosetta™ 2 (DE3) competent cells (Novagen, 71397, from Sigma-Aldrich) via chemical transformation as described in section 2.1. Transformed Rosetta cells were plated on the selective LB agar plates containing chloramphenicol (35 ug/ml final concentration) and additional antibiotics corresponding to selective markers of the carried plasmids. Colonies were picked the next day, seeded in a 20 ml selective LB liquid

culture with the same antibiotic combinations, and incubated in a 37 ℃ shaker overnight. The 20 ml seeding culture was transferred to a fresh 1L LB culture with the same antibiotic combination and induced at O.D. around 0.6-0.8. Codon-optimized protein expression cell culture was induced by various amounts of isopropyl β-D-1-thiogalactopyranoside (IPTG) at 16 ℃ for 16-20 hours. Cell culture for viral RNA expression was induced by different concentrations of IPTG at 16 ℃ for 16-20 hours, at 25 ℃ for 9-12 hours, at 28 ℃ for 8-9 hours, at 30 ℃ for 6-8 hours, or at 37 ℃ for 4-6 hours. The expression culture with pBAD18 vectors was induced by 0.01% L-arabinose (final concentration) at O.D. 0.1-0.2. The following tables summarize all the induction conditions tested in this work.

Expression Construct	IPTG/Arabinose Final Concentration	Induction Temperature and Duration	Plasmids
ORF2	0.75 mM IPTG	16 ℃ for 16-20 hours 37 ℃ for 4-6 hours	Table 2.1.3
hPBV1 and ORF2 co-expression from a pETDuet plasmid	1, 10, 100, 750 μM IPTG	16 ℃ for 16-20 hours	Table 2.1.6
hPBV1 and ORF2 co-expression from a pETDuet plasmid	0.5 mM IPTG	37 ℃ for 4-6 hours	Table 2.1.6
CP/ΔCP and ORF2 co-expression	0.01 or 0.5 mM IPTG	16 ℃ for 16-20 hours	Table 2.1.4

Table 2.2.1: Summary of induction conditions for ORF2 related expression. ORF2 refers to the full-length ORF2 and its various truncated mutants in this table.

Expression Construct	IPTG/Arabinose Final Concentration	Induction Temperature and Duration	Plasmids
Codon-optimized CP and/or RdRP	0.75 mM IPTG	16 ℃ for 16-20 hours	Table 2.1.1 and 2.1.2
hPBV1 expression	0.75 mM IPTG	16 ℃ for 16-20 hours 37 ℃ for 4-6 hours	Table 2.1.5
hPBV2 expression	0.75 mM IPTG	16 ℃ for 16-20 hours 25 ℃ for 9-12 hours 37 ℃ for 4-6 hours	Table 2.1.7
pBAD18-KmR-His-RdRP	0.01% L-Arabinose	16 ℃ for 16-20 hours	Table 2.1.2

Table 2.2.2: Summary of induction conditions for CP, RdRP, hPBV1, and hPBV2 separate expression.

Co-expression Construct	IPTG/Arabinose Final Concentration	Induction Temperature and Duration	Plasmids
hPBV1 and RdRP from pETDuet	0.75 mM IPTG	16 °C for 16-20 hours	Table 2.1.8
Codon-optimized CP and hPBV2	0.75 mM IPTG	16 °C for 16-20 hours	Table 2.1.1 and 2.1.7
hPBV1 and hPBV2 from separate plasmids	1 mM IPTG	25 °C for 9-12 hours 37 °C for 4-6 hours	Table 2.1.5 and 2.1.7
hPBV1 and hPBV2 from separate plasmids	0.01, 1 mM IPTG	16 °C for 16-20 hours	Table 2.1.5 and 2.1.7
hPBV1 and hPBV2 from pETDuet	0, 0.01, 0.1, 1 mM IPTG	16 °C for 16-20 hours	Table 2.1.8
hPBV1 and hPBV2 from pETDuet	0, 0.01, 0.1, 1 mM IPTG	37 °C for 4-6 hours	Table 2.1.8
hPBV1 and hPBV2 from pETDuet	1 mM IPTG	25 °C for 9-12 hours 28 °C for 8-9 hours 30 °C for 6-8 hours	Table 2.1.8
hPBV1 and hPBV2 (GDD-GAA mutant)	1 mM IPTG	25 °C for 9-12 hours	Table 2.1.8
pETDuet-ΔRBS-hPBV and pET28b-His-RdRP	1 mM IPTG	16 °C for 16-20 hours	Table 2.1.8 and 2.1.2
pETDuet-ΔRBS-hPBV and pBAD18-KmR-His-RdRP	0.01% L-arabinose 1mM IPTG	Keep in 37 °C after arabinose induction and move to 16 °C after IPTG induction for 16-20 hours	Table 2.1.8 and 2.1.2

Table 2.2.3: Summary of induction conditions for hPBV1 and hPBV2 co-expression.

Co-expression Construct	IPTG/Arabinose Final Concentration	Induction Temperature and Duration	Plasmids
hPBV1 and hPBV2 (GFP-tagged RdRP)	0.01, 0.1, 1 mM IPTG	16 °C for 16-20 hours	Table 2.1.1 and 2.1.7
CP/hPBV1 and GFP-tagged RdRP	0.75 mM IPTG	16 °C for 16-20 hours	Table 2.1.1, 2.1.2, and 2.1.5
CP and GFP	0.75 mM IPTG	16 °C for 16-20 hours	Table 2.1.1 and 2.1.2

Table 2.2.4: Summary of induction conditions for the co-expression of CP and GFP-tagged RdRP.

2.3. Virus-like particle (VLP) purification

E. coli Rosetta cells were pelleted from 1L culture by 4,500g centrifugation (Beckman Coulter) for 15 minutes. The harvested cell pellet was resuspended in or in lysis buffer (300 mM NaCl, 50 mM Tris, pH 7.5, 10 % glycerol). Cells were lysed by three cycles of a 5-minute sonication program (Branson Sonifier 250) with the following parameters: 1 sec ON and 1 sec OFF, 70% amplification. The insoluble fraction was removed by 20,000g centrifugation for 45 minutes using a JA-25.50 rotor. Recombinant VLPs remained in the supernatant (i.e., soluble fraction) and were then pelleted by Beckman Coulter L80XP ultracentrifuge at 37500 rpm using an SW41-Ti rotor. The VLP pellet was resuspended and loaded to the top of a pre-loaded CsCl gradient (1 ml of 1.45 g/cm^3 CsCl, 3 ml of 1.40 g/cm^3 CsCl, 2 ml of 1.275 g/cm^3 CsCl, and 2 ml of 1.25 g/cm^3 CsCl). After spinning at 37000 rpm with the SW41-Ti rotor at 4 °C for 18-24 hours, VLPs eventually appeared as a white light-scattering band around 1.3-1.35 g/cm^3 CsCl density. The VLP fraction was collected by puncturing the Ultra-Clear tubes (Beckman Coulter 344059) at the band location with a 3-ml syringe. The VLP fraction from CsCl density gradient ultracentrifugation was buffer exchanged to remove CsCl. For further purification, the VLP sample was mixed with 60% iodixanol (OptiPrep™ Density Gradient Medium) to a final concentration of 40% iodixanol. The VLP-mixed 40% iodixanol was separated by self-generated gradient ultracentrifugation with a Type 80 Ti rotor at 60,000 rpm and 4 °C for 16 hours. The purified VLPs showed up as a diffuse band at around 40% iodixanol density and were collected by fractionating from top to bottom.

2.4. Immunoprecipitation

Around 0.2 mg of the purified VLPs from ultracentrifugation were mixed with 25 μl of Pierce™ protein A/G magnetic beads (Thermo Scientific™ 88802) and 10 μl of the anti-PBV CP polysera (from Pacific Immunology, Ramona, CA, US) in 500 μl lysis

buffer or PBS buffer. The mixture was rotated at 4 ℃ for overnight. The immunoprecipitated VLPs were separated from eluants in the mixed solution using a magnetic stand, which thoroughly attracted all magnetic beads to the side of the tube. The supernatant containing unbounded RNA and proteins was gently removed and collected in a separate microcentrifuge tube as the corresponding flow-through sample. The magnetic agarose beads were gently washed thrice with 500 μl PBS buffer. After the removal of the third wash, all proteins bound to the beads were denatured in 50 μl PBS buffer with 10 μl of the 6× SDS-PAGE sample loading dye and directly boiled at 100 ℃ for 3 minutes. The immunoprecipitated protein samples were analyzed by SDS-PAGE, Western blots, and MS sequencing. The immunoprecipitated RNA was extracted as described in section 2.5.

2.5. RNA extraction and cDNA synthesis

Each 100 μl of the test sample was mixed with 300 μl of the Tri Reagent® from Zymo Research. Total RNAs were extracted by the Direct-zol RNA Microprep kit (Zymo Research R2061) with an on-column DNase I digestion to remove contaminated DNAs and eluted in 15 μl of DEPC nuclease-free water. RNA concentration was measured by Thermo Scientific™ NanoDrop™ 2000. For cDNA synthesis, 5.5 μl of DEPC water containing around 2000 ng of total RNA sample was mixed with 0.5 μl of random hexamers and 0.5 μl of 10 mM dNTP mix and then heated at 65 ℃ for 5 minutes in a thermal cycler (Bio-Rad T100™). Tagged gene-specific reverse primers replaced random hexamers in strand-specific RT-qPCR (section 2.6). The pre-heated mixture was immediately chilled on ice for 2 minutes before mixing with 2 μl of 5× SSIV Buffer, 0.5 μl of 0.1 M dithiothreitol (DTT), 0.5 μl of ribonuclease inhibitor, and 0.5 μl of SuperScript™ IV reverse transcriptase (200U/ml) (Invitrogen) to a total volume of 10 μl. The standard reverse transcription (RT) reactions with random hexamers were incubated at 23 ℃ for 10 minutes, extended at 50 ℃ for 20 minutes, inactivated at 80 ℃ for 10 minutes, and finally cooled down to 4 ℃. The strand-specific RT reactions with tagged primers were directly extended at 50-55 ℃ for 20 minutes, inactivated at 80 ℃ for 10 minutes, and cooled down to 4 ℃.

2.6. Strand-specific RT-qPCR

The predicted secondary structure of hPBV UTRs at 5' and 3' extremes indicated potential self-priming that could result in primer-independent cDNA synthesis and a biased outcome. Therefore, strand-specific RT-qPCR was required to specifically detect cDNA amplified by the designated primers and to differentiate the positive and negative strands at the same time[84,85]. In principle, strand-specific RT-qPCR relies on tagged primers to add a barcode sequence at the 5' end of the synthesized cDNA. During real-time PCR, the barcoded cDNA was detected by a forward primer as the barcode sequence and a strand-specific reverse primer of the targeted viral RNA. The barcode sequences for viral RNA detection were 5'-GACCGTCATGATGGCGAATA for positive-sense viral RNA and 5'-GCTAGCTTCAGCTAGGCATC for negative-sense viral RNA. The barcode sequence for HDV ribozyme sequence and viral RNA 3' end detection primers was 5'-GCTAGCTTCAGCTAGGCATC. The detailed gene-specific primer sequences were listed in Table 2.6.

Each 10 µl qPCR reaction consisted of 1 µl of ten-fold diluted cDNA, 0.8 µl of the primer pair 1:1 mix (25 µM), 3.2 µl of nuclease-free water, and 5 µl of 2× PerfeCTa SYBR® Green SuperMix (Quantabio). Real-time PCR reactions were run by a Bio-Rad C1000™ touch thermal cycler with CFX96™ optical reaction module (i.e., a real-time PCR system) for 40 cycles. The real-time PCR cycle parameters were 95 ℃ for 3 minutes, followed by 40 cycles of 95 ℃ denaturation for 10 seconds and 60 ℃ extension for 30 seconds. Fluorescent signal readings for SYBR were recorded at the end of each cycle. After real-time PCR reactions, a melting temperature measurement from 65 to 95 ℃ was carried out to validate the primer specificity.

Target		Purpose	Sequence
rrsA	Forward	Real-time PCR	CTCTTGCCATCGGATGTGCCCA
	Reverse		CCAGTGTGGCTGGTCATCCTCTCA
idnT	Forward	Real-time PCR	CTGTTTAGCGAAGAGGAGATGC
	Reverse		ACAAACGGCGGCGATAGC
ihfB	Forward	Real-time PCR	GCGGTTTCGGCAGTTTCT
	Reverse		CGCAGTTCTTTACCAGGTTT
hPBV1	Positive strand	Reverse transcription	**5'-GACCGTCATGATGGCGAATA** AAAGGGACCAAACGACTCCT
	Forward	real-time PCR	GACCGTCATGATGGCGAATA
	Reverse		AAAGCAACTCCACACGACCT
hPBV1	Negative strand	Reverse transcription	**5'-GCTAGCTTCAGCTAGGCATC** CAGCCGACACACTAACATC
	Forward	real-time PCR	GCTAGCTTCAGCTAGGCATC
	Reverse		CGACATTGTGACGAGACGTT
hPBV2	Positive strand	Reverse transcription	**5'-GACCGTCATGATGGCGAATA** CATACGTCTGGGTCATAGTAGC
	Forward	real-time PCR	GACCGTCATGATGGCGAATA
	Reverse		CGATTAGCCACCATACCACTC
hPBV2	Negative strand	Reverse transcription	**5'-GCTAGCTTCAGCTAGGCATC** GGTATGGTATCCATGGAATGGG
	Forward	real-time PCR	GCTAGCTTCAGCTAGGCATC
	Reverse		TGATGCCACCTGCGTAAAT
GFP	Reverse	Reverse transcription	**5'-GACCGTCATGATGGCGAATA** CAAGACTGGACCATCACCAA
	Forward	real-time PCR	GACCGTCATGATGGCGAATA
	Reverse		CAAGACTGGACCATCACCAA
hPBV1	HDV	Reverse transcription	**5'-GCTAGCTTCAGCTAGGCATC** CGCGAGGAGGTGGAGAT
	3' end	Reverse transcription	**5'-GCTAGCTTCAGCTAGGCATC** GAGGTGTGTGCTTAACGCA
	Forward	real-time PCR	GCTAGCTTCAGCTAGGCATC
	Reverse		GCTCAGCTGAATGATTGGACC
hPBV2	HDV	Reverse transcription	**5'-GCTAGCTTCAGCTAGGCATC** CGCGAGGAGGTGGAGAT
	3' end	Reverse transcription	**5'-GCTAGCTTCAGCTAGGCATC** GCAGTTGGGACTGTTAGTCC
	Forward	real-time PCR	GCTAGCTTCAGCTAGGCATC
	Reverse		CCTTGGCTACACCAAGTCG

Table 2.6: Primers for standard and strand-specific RT-qPCR. The barcode sequences in gene-specific reverse primers are bolded and started from the 5'-end.

2.7. Western blot and quantification

After separating the protein samples on SDS-PAGE, the resolving gel and a methanol-soaked 0.2 μm PVDF membrane (Thermo Scientific™ 88520) were sandwiched between two thick filter sponges. The sandwich was tightly squeezed to remove air bubbles and extra liquid before being inserted into the transfer apparatus. The PVDF membrane was placed near the cathode while the gel was connected to the anode side. Western blots were transferred in Tris-glycine transfer buffer at 100 V at 4 °C for 60 minutes. The transferred PVDF membrane was blocked in 10 ml of TBST buffer with 5% blotting-grade blocker (Bio-Rad 1706404) for 1 hour at room temperature and followed by three gentle washes in TBST buffer. The membrane was then incubated overnight at 4 °C in 10 ml TBST buffer with 2 μl commercial primary antibodies or 10 μl of anti-PBV CP polysera (raised in rabbit, from Pacific Immunology, Ramona, CA, US). Commercial primary antibodies used in this work were anti-His antibodies (mouse, Invitrogen MA1-135), anti-Strep antibodies (mouse, iba 2-1509-001), GAPDH loading control antibodies (mouse, Invitrogen MA5-15738), anti-GFP antibodies with Alex fluor 488 (mouse, R&D Systems™ IC42402G), anti-His antibodies with Alex fluor 488 (mouse, Invitrogen, MA1-21315-A488). After three washes in TBST buffer, membranes incubated Alex-fluor-488-conjugated primary antibodies were directly visualized under Sapphire RGBNIR Biomolecular Imager (Azure Biosystems, Inc) in the Alex 488 channel. Anti-PBV CP polysera treated membranes were incubated in 10 ml TBST buffer with 2 μl of goat-anti-rabbit IgG secondary antibody (Invitrogen 31340) for 1 hour at room temperature. Membranes incubated with the other primary antibodies were soaked in 10 ml TBST buffer containing 2 μl of goat-anti-mouse IgG secondary antibody (Invitrogen 31322) for 1 hour. Membranes treated with secondary antibodies were washed three times in TBST buffer before staining in BCIP®/NBT alkaline phosphatase substrate solution (Sigma-Aldrich SIGMA*FAST*™, B5655). The staining solution was prepared by dissolving one tablet in 10 ml Milli-Q water. After staining, membranes were rinsed in excess Milli-Q water to stop exposure and dried in a chemical hood. Western blots were imaged by Epson Perfection V600 Photo Scanner and saved in .tif format. Western blot

signals were quantified by ImageJ. The noise was subtracted by measuring a same-sized area from the background region on the same image.

2.8. Transmission electron microscopy (TEM)

VLP samples were stained with uranyl formate on Formvar/Carbon 400 mesh grids (Ted Pella, Inc, 01754-F). The filtered 0.75% uranyl formate solution (Polysciences, 24762-1) was freshly prepared in boiled Milli-Q water and titrated with 5 M NaOH to a slightly orange color (i.e., 0.75 mg uranyl formate per 1 ml water, approximately 2 μl of 5 M NaOH titrated per 1 ml of uranyl formate). Carbon-coated grids were glow discharged at 5 mA for 1 minute and then loaded with 3.5 μl of diluted VLP sample (< 0.05 mg/ml) for 60 seconds. After two sequential washes in the Milli-Q water droplets and one wash in a 0.75% uranyl formate droplet, the grid was stained in a 0.75% uranyl formate droplet for 90 seconds. Each droplet was 25 μl in volume and placed on a parafilm. The residual staining solution on the grid was absorbed by filter paper. Grids were imaged by a JEOL JEM-1400 Flash Transmission Electron Microscope, equipped with a high-contrast polepiece and an AMT NanoSprint15 sCMOS camera. All waste materials containing uranyl formate were discarded separately and collected by a designated person from Rice EHS.

2.9. Nanosight

The Nanosight NS300 was used to visualize VLP samples for particle size measurement and track the RdRP-packaged VLPs by the green fluorescence from the GFP-tagged RdRP. Before imaging, Nanosight tubing was carefully washed with 70% ethanol and sufficient Milli-Q water to remove all visible bubbles. The camera and the chip were cleaned with a Kim-wipe with water if needed. Before sample loading, the tubing inside the imaging chamber was rinsed with sufficient PBS buffer to remove residual particles in the imaging area. All washes and samples were injected by a syringe pump. The particle concentration of purified VLPs was estimated from the protein

concentration measured by Bradford assays. GFP-tagged viral particles were diluted to 10^7-10^9 particles/ml and tracked under Blue Laser 488 nm under Nanosight. Samples were measured by real-time video recording for 1 minute with a syringe pump speed of 80-100 μl/min. The system-programmed software utilized light scattering from particles and Brownian motion to calculate the size distribution of particles by tracking their movements in liquid suspension.

2.10. Thermal shift assay

In conventional thermal shift assays, exposed hydrophobic surfaces interact with dye during protein unfolding, increasing fluorescence signal readings. Thermal shift assays with fluorophores were usually used to measure protein melting temperature, identified as the temperature at which half the protein is denatured. However, a GFP-tagged protein had its inherent fluorescent signals from GFP. With increasing temperatures, GFP fluorescence signals decreased as the protein unfolded. Thermal shift assays for GFP-tagged proteins could be carried out without additional fluorophores to measure the relative abundance of GFP-tagged proteins in a mixed protein sample like VLPs.

The purified VLP sample containing GFP-tagged RdRP was concentrated to around 0.5 mg/ml. A serial VLP dilution was prepared in one-fold, two-fold, and four-fold dilutions. Positive controls were a series of GFP dilutions (1× and 10×) and GFP-CP dilutions (1×, 2×, and 4×). Each 10 μl of various diluted sample were loaded on a 96-well plate. The thermal shift assay was measured by a Bio-Rad C1000™ touch thermal cycler with CFX96™ optical reaction module. Fluorescence signals were measured with increasing temperatures from 20 ℃ to 100 ℃.

2.11. Circular dichroism (CD)

CD measurement was used in this work to characterize the secondary structure changes of an interested protein in a temperature scanning mode. The protein

concentration of the purified interested protein was measured by Bradford assay using UV/Vis spectrometer at 595 nm. The protein sample was diluted to 0.5 μM in PBS buffer or other tested buffers. A special 1-cm length quartz cuvette was washed with detergent and rinsed with sufficient 70% ethanol. After completely drying with air gas, the cuvette was loaded with a blank (buffer without any protein) for blank measurement and recorded in the system. After removing the buffer, the cuvette was loaded with 500 μl of the diluted 0.5 μM protein sample. The CD spectra were measured in a temperature scanning mode from 15 °C to 85 °C with 10 °C intervals for wavelengths from 195 nm to 255 nm with 2 nm intervals.

Chapter 3

Plasmid-based expression of the recombinant hPBV

According to the life cycle of a typical dsRNA (Fig. 1.1), +ssRNA is responsible for the translation of all viral proteins and serves as the template for RNA replication after viral assembly. Therefore, the key strategy of my recombinant hPBV expression system is to express the exact viral +ssRNA transcripts from plasmids in *Escherichia coli* Rosetta 2 (DE3) cells (Fig. 3). The cDNA of each hPBV RNA segment is flanked by the T7 promoter sequence and the antigenomic hepatitis delta virus (HDV) ribozyme sequence at the 5' and 3' extremes, respectively[86,87]. *E. coli* Rosetta 2 cells contain a pLysS plasmid to express eukaryotic tRNAs (AGG, AGA, AUA, CUA, CCC, GGA, and CGG) rarely used in *E. coli* to overcome potential codon bias for heterogenous protein expression[88,89]. The DE3 strain of *E. coli* Rosetta 2 cells contains a T7 RNA polymerase gene governed by an IPTG-inducible lacUV5 promoter. Therefore, IPTG induction allows T7 RNA polymerase to transcribe the designed sequence downstream of a T7 promoter. After IPTG induction, T7 RNA polymerase starts transcription synthesis from the guanine (G) after the T7 promoter sequence, which ends in TATA. Coincidentally, the 5'-termini of two hPBV +ssRNA segments start with GTAAA. Thus, viral RNA transcription starts exactly after the T7 promoter. As RNA transcription terminates, the HDV ribozyme sequence cleaves itself at the 3' end, releasing viral RNA segments with the exact 5'- and 3'-extremes. Suppose the predicted RBSs on viral RNA are functional. In that case, the inherent RBSs recruit bacterial ribosomes and initiate protein translation (Fig. 1.2.2). Viral protein expression can be monitored by Western blots to conclude if

the predicted RBSs in the hPBV genome are fully functional. The viral +ssRNA and proteins would eventually assemble into nascent viral particles, i.e., recombinant hPBVs.
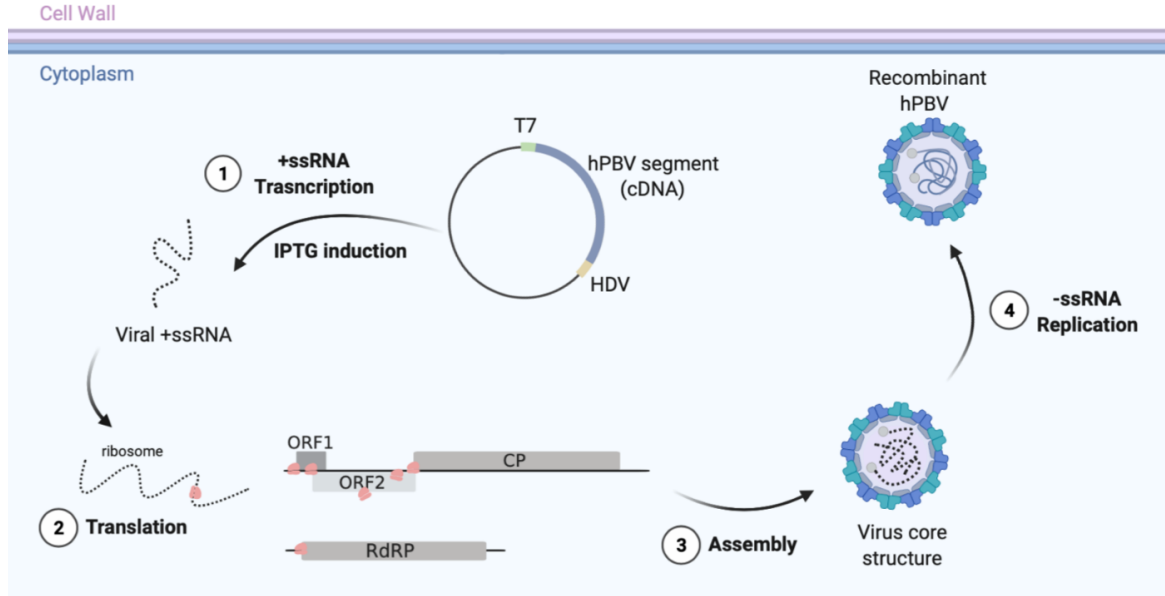
Figure 3: Scheme of viral RNA segment expression in *E. coli* Rosetta 2 (DE3) cells.

3.1. Characterize the expression of hPBV RNA segment 1

The viral RNA segment 1 of hPBV (hPBV1) was expressed from a pET28a plasmid. The cDNA of hPBV1 was inserted right after the backbone T7 promoter, replacing the Lac operator (LacO) and RBS sequences on the pET28a vector (Fig. 3.1.1*A*). Therefore, the expression of a viral protein implies the proper functionality of its upstream predicted RBSs. Among the four predicted viral proteins in the hPBV genome, CP encoded by hPBV1 is the only structural protein of hPBV capsids and automatically assembles into virus-like particles (VLPs)[37,83]. Therefore, the successful expression of CP from the viral RNA expression of hPBV1 builds the foundation for generating recombinant hPBVs in an *E. coli* expression system. For detection purposes, CP was fused with an N-terminal His or Strep tag in pET28-ΔRBS-hPBV1.3 and pET28-ΔRBS-hPBV1.4, respectively (Fig. 3.1.1*B*). The expression of CP from the two plasmids was monitored by Western blots with and without IPTG induction at 16 °C and 37 °C. A higher expression level of CP was observed with IPTG induction, indicating that the

upstream predicted RBS of CP was actively functional. To determine whether the expressed CP from hPBV1 expression was folded correctly and behaved similarly to previous reports, I purified the assembled VLPs from hPBV1 expression in comparison with the VLPs from codon-optimized CP expression. After CsCl density gradient ultracentrifugation, they both showed up as a white band with light scattering at around 1.30-1.32 g/cm^3, indicating they formed virus-like particles (Fig. 3.1.2*A*). The VLPs from hPBV1 expression penetrated at a similar density as the VLPs from the codon-optimized CP (by pET19b-CP). VLPs from hPBV1 expression and codon-optimized CP expression showed a major protein band on SDS-PAGE at a nearly identical position, consistent with the calculated molecular weight of CP (~ 62.0 kDa) (Fig. 3.1.2*B*). The identity of CP protein was further confirmed by Western blots, where the ~ 62.0 kDa protein band was recognized by anti-PBV CP polysera (from Pacific Immunology, Ramona, CA, US) (Fig. 3.1.2*C*)[83]. The results suggested that CP from hPBV1 expression could automatically assemble into VLPs, behaving similarly to the CP from codon-optimized protein expression. The VLP production from hPBV1 expression built the foundation of the recombinant hPBV expression system.
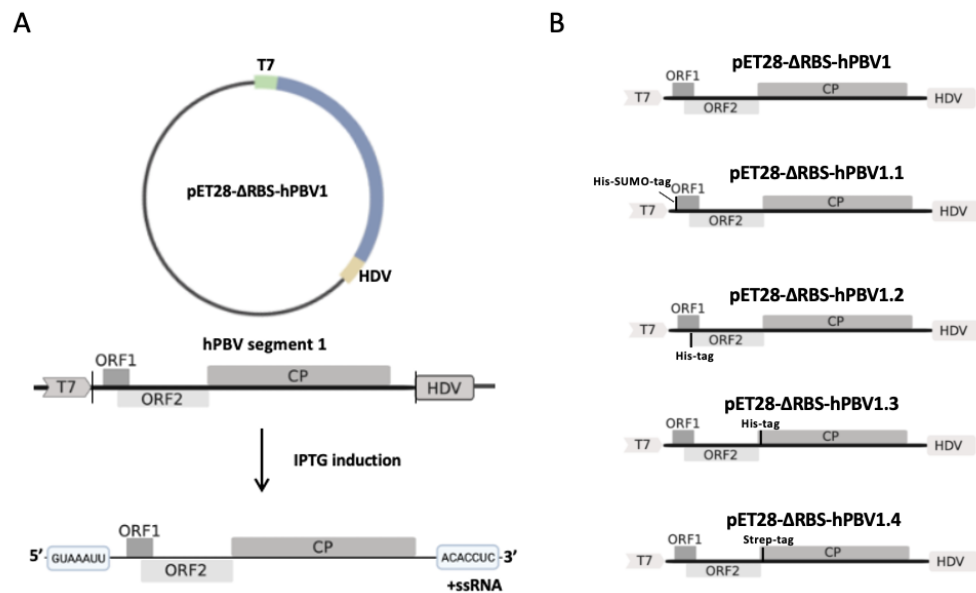
Figure 3.1.1: Scheme of for hPBV RNA segment 1 expression. (A) The expression scheme details for hPBV1 RNA segment expression. (B) Different plasmid constructs of hPBV1 used in this work for each ORF protein expression detection.

According to dsRNA viral replication cycle, the viral mRNA, i.e., the +ssRNA transcript, encodes all viral proteins required for viral assembly. To detect the expression of rest predicted viral proteins on hPBV1, the 39 amino-acid ORF1 was fused with a His-SUMO tag as His-SUMO-ORF1 (~19 kDa), while the 224-amino acid ORF2 was labeled with a His tag as a His-ORF2 (~25 kDa) (Fig. 3.1.1*B*). Fusion tags were inserted into the N-termini of viral proteins, i.e., at the start of coding regions, which would not interrupt the predicted inherent RBSs. For instance, a C-terminal fusion tag on ORF2 would interfere with the expression of CP from its upstream inherent RBS, as ORF2 and CP coding sequences overlapped their termination and initiation codons in a -1 frameshift (Fig. 1.2.3.1). His-SUMO-ORF1 and His-ORF2 were expressed by engineered hPBV1 from pET28-ΔRBS-hPBV1.1 and pET28-ΔRBS-hPBV1.2, respectively. The engineered viral expression was induced by 1 mM IPTG induction at 16 °C for 20 hours. Since ORF1 and ORF2 proteins were functionally uncharacterized, His-SUMO-ORF1 and His-ORF2 were purified by Ni-NTA resins as His-tag fusion proteins. Western blots against anti-His antibodies showed that His-SUMO-ORF1 was soluble and bound to Ni-NTA resin (Fig. 3.1.2*C*). However, the His-tagged ORF2 was not detected in the Ni-NTA
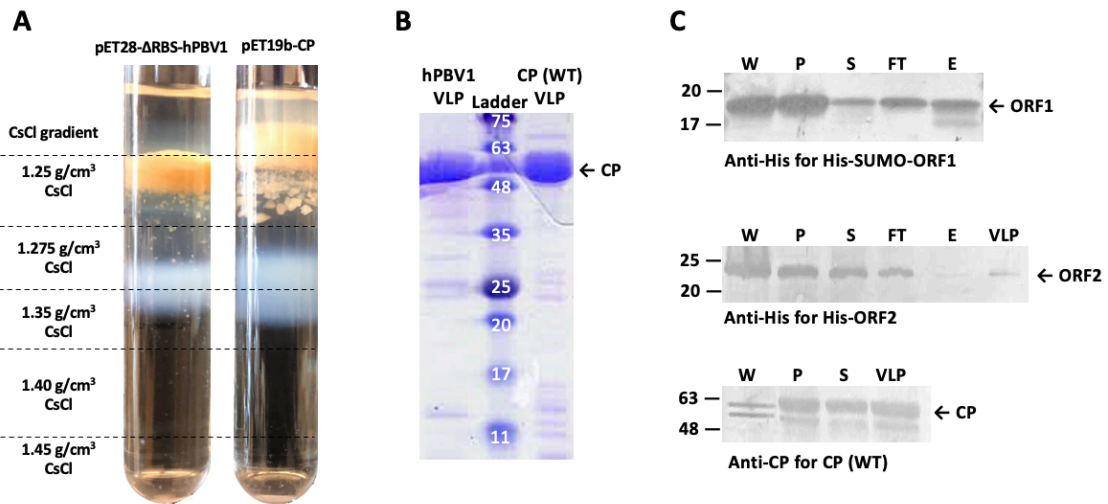
Figure 3.1.2: Results for hPBV RNA segment 1 expression. (A) Ultracentrifuge results for hPBV1 and CP expression comparison. (B) SDS-PAGE to compare the protein profile of VLPs yielded from hPBV1 and CP expression. (C) Western blot results for each ORF detection on hPBV1. W: whole lysate; P: insoluble pellet; S: soluble supernatant; FT: flow-through; E: elution from Ni-NTA gravity column; VLP: purified VLP fraction from CsCl density gradient ultracentrifugation.

elution fraction. Since His-ORF2 were expressed but not bound to Ni-NTA resins, I loaded the Ni-NTA flowthrough fraction to CsCl density gradient ultracentrifugation for VLP purification and found His-ORF2 in the purified VLP fractions (Fig. 3.1.2*C*). Detection of His-ORF2 in VLP fractions explained that the N-terminal His tag of ORF2 was hindered from binding to Ni-NTA resin due to its interaction with CP or packaging in assembled particles.

3.2. Characterize the expression of hPBV RNA segment 2

The expression of hPBV viral RNA segment 2 (hPBV2) was tested on a pEXT20 plasmid. The pEXT20 plasmid expresses proteins under a pTac promoter. The pEXT20-T7-hPBV2 plasmid replaced the pTac promoter in the pEXT20 backbone with the T7-hPBV2-HDV cassette (Fig. 3.2*A*). Like the hPBV1 expression scheme, the T7 RNA polymerase synthesizes RNA transcripts with the exact 5'-end of viral hPBV2 after IPTG
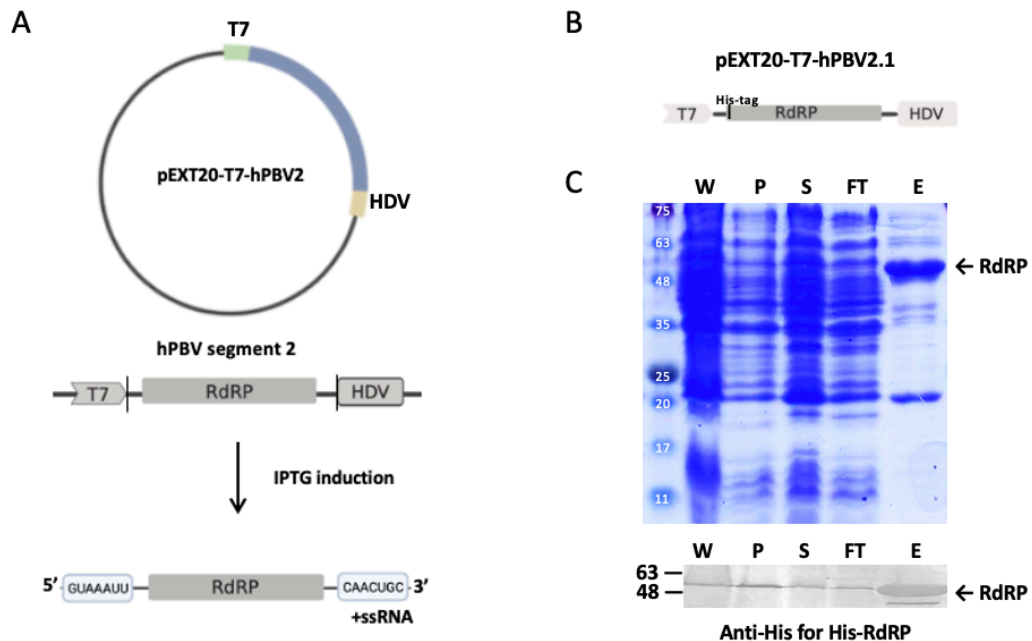
Figure 3.2: Scheme and results for hPBV RNA segment 2 expression. (A) Viral RNA segment 2 expression details after IPTG induction. (B) Plasmid construct used for RdRP detection. (C) SDS-PAGE and Western blot results for RdRP detection from hPBV2 expression. W: whole lysate; P: insoluble pellet; S: supernatant; FT: flowthrough; E: elution from Ni-NTA gravity column.

induction. At transcription termination, the HDV ribozyme sequence at the 3'-end cleaves itself and release an RNA transcript with the exact 5'- and 3'-termini of viral hPBV2. To detect the expression of RdRP from hPBV2, pEXT20-T7-hPBV2.1 transcribed an engineered hPBV2 which encoded an N-terminal His-tagged RdRP (Fig. 3.2*B*). His-RdRP expressed from the engineered hPBV2 was first detected by anti-His Western blots with and without a 6-hour IPTG induction at 37 °C. His-RdRP expression was only detected under IPTG induction, indicating the successful integration of the T7 promoter in the pEXT20 vector, which regulated the RNA transcription under IPTG induction. According to the dsRNA virus life cycle, viral RdRP should be fully activated upon signals from capsid proteins after packaging into nascent viral particles. The His-tagged RdRP was characterized by its binding affinity to Ni-NTA resins. As SDS-PAGE and Western blots presented, His-RdRP translated from the engineered hPBV2 was successfully purified by a Ni-NTA gravity column as a protein at ~ 60 kDa (Fig. 3.2*C*). The results implied that the upstream inherent RBS of RdRP on hPBV2 actively recruited bacterial ribosomes for protein translation.

3.3. Characterize the co-expression of both viral RNA segments

As illustrated from the *in situ* transcribing structures of turreted reoviruses and rotaviruses, each encapsulated RdRP or TECs held the terminus of one genomic RNA segment at a five-fold vertex[29,31,32,40]. With up to 12 packaged RNA segments, the +ssRNA transcript of each genomic dsRNA segment was synchronously synthesized and released after viral infection. Though the situation was complicated for bacteriophage φ6, which envelops various copy numbers of RdRP, 3-4 copies of RdRP were optimal for RNA transcription synthesis, indicating approximately one RdRP per RNA segment[90]. In general, each viral RNA segment was believed to be transcribed by one designated RdRP, resulting in similar concentrations of transcripts from each viral RNA segment. To mimic viral RNA transcription from a viral infection, the viral RNA transcripts of hPBV1 and hPBV2 were co-expressed from a pETDuet plasmid, allowing simultaneous transcription of hPBV1 and hPBV2 after two T7 promoters (Fig. 3.3*A*). In other words, the upstream T7 promoter of each viral RNA expression cassette located T7 RNA

polymerase for hPBV1 or hPBV2 transcription as a designated RdRP for each viral RNA segment in a transcribing virion. Since all predicted ORFs were successfully detected after the expression of individual viral RNA segments, I expected to purify recombinant hPBVs resembling the isolated PBVs from feces samples.

The co-expression was first optimized under various induction conditions. Parameters for the tested conditions included: 16 °C induction for 20 hours at 0.001, 0.01, 0.1, and 1 mM IPTG; 37 °C induction for 6 hours at 0.001, 0.01, 0.1, and 1 mM IPTG; and 1 mM IPTG induction at 25 °C for 12 hours, 28 °C for 9 hours, and 30 °C for 8 hours. Under each induction condition, hPBV1 expression was characterized by CP as VLP yield, while His-RdRP yield from Ni-NTA purification implied hPBV2 expression. The optimal co-expression condition was 1 mM IPTG induction at 25 °C for 12 hours, which exhibited a decent yield of VLP and a detectable amount of RdRP in concentrated
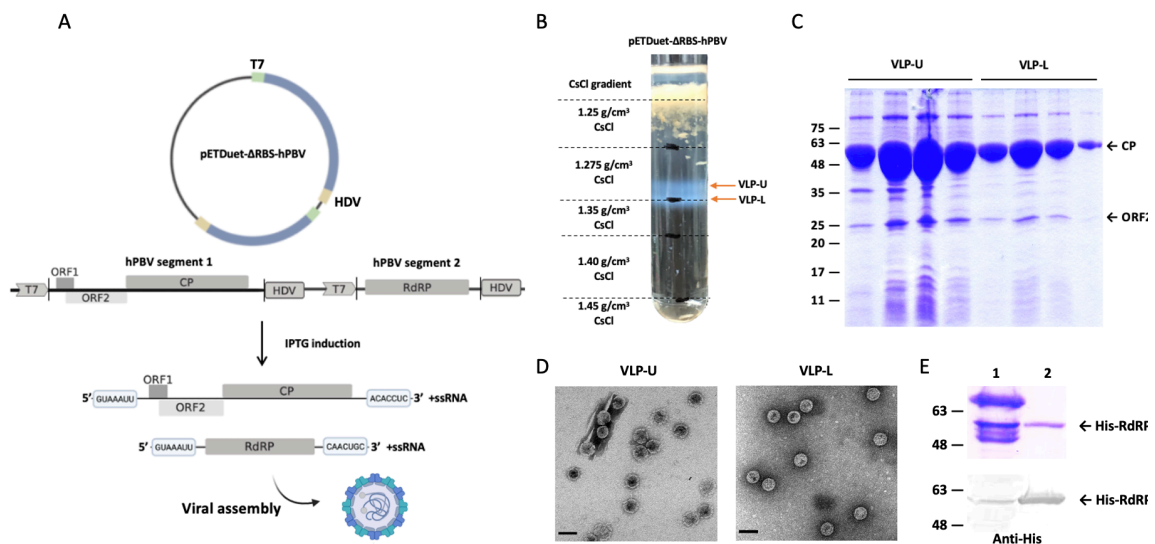
Figure 3.3: Scheme and results for the co-expression of both viral RNA segments.
(A) Expression details of both viral RNA segments after IPTG induction. (B) VLP purification results for co-expression (C) SDS-PAGE results for the purified VLPs after the iodixanol gradient ultracentrifugation. (D) TEM images for VLP-U and VLP-L morphology characterization. (E) SDS-PAGE and Western blot for RdRP expression detection in co-expression vs. single expression. Lane 1: concentrated Ni-NTA elution E1 fraction from pETDuet-ΔRBS-hPBV.1 expression (Table 2.1.5 and Fig. 3.5.1*B*). Lane 2: Ni-NTA elution E1 fraction from pEXT20-T7-hPBV2.1 expression.

Ni-NTA elution (Fig. 3.*3B* and *E*). After CsCl gradient ultracentrifugation, two closely spaced VLP bands were observed at approximately 1.33-1.35 g/cm^3 density (Fig. 3.3*B*).

The relatively upper and lower VLP layers (i.e., VLP-U and VLP-L) were collected separately and loaded to an iodixanol density gradient ultracentrifugation. VLP-U and VLP-L fractions appeared at around 40% iodixanol density ($\sim$ 1.22 g/cm^3), a density usually expected for recombinant viruses at a similar size, like recombinant Adeno-associated viruses (AAVs)[91]. The VLPs purified by iodixanol density gradient ultracentrifugation were fractionated and collected from top to bottom. According to the SDS-PAGE analysis of fractions from VLP-U and VLP-L iodixanol density gradient ultracentrifugation, two dominant bands at approximately 62 kDa and 25 kDa correspondingly to the expected molecular weights for CP and ORF2, respectively, were clearly observed in all VLP fractions (Fig. 3.3*C*). Central fractions of VLP-L and VLP-U were analyzed under TEM with negative staining (Fig. 3.3*D*). VLPs from VLP-L and VLP-U were observed as round particles with a diameter of around 35 nm. VLP-U particles often appeared partially broken, with the interior filled with stains. In contrast, VLP-L was primarily intact with stain excluded, where most of the particle interior was occupied with some internal densities, which would potentially be packaged proteins. Thus, VLP-L was further analyzed by protein identity mass spectrometry (MS) sequencing (UTHealth Proteomics Service Center). The MS sequencing results identified the viral proteins ORF2 and CP as the top two matches, consistent with the $\sim$ 60 kDa and $\sim$ 25 kDa bands observed on SDS-PAGE (Table 3.3). The rest hits were contaminated ribosomal proteins due to non-specific packaging via the interaction between the positively-charged interior-facing N-terminal domain of CP and negatively charged ribosomal RNAs (rRNAs) incorporated in ribosomal proteins[37]. However, no peptide match was found for ORF1, indicating that ORF1 was probably not an encapsulated viral protein. Unexpected, viral RdRP, which played an essential role in transcription and replication inside the virus particles, was not detected in the purified recombinant hPBV VLPs by MS sequencing. To confirm the RdRP detection results in recombinant VLPs, I attempted the co-expression of the two viral RNA segments by separate plasmids, pET28-ΔRBS-hPBV1 and pEXT20-T7-hPBV2.1. However, viral RdRP was still not detectable in the purified recombinant VLPs. Missing RdRP in the recombinant hPBVs hinted that the recombinant particles were not infectious and explained that no or little *E. coli* host cells were lysed during the recombinant hPBV expression.

Since unpackaged RdRP was detectable in the elution from the Ni-NTA column but no packaged RdRP was detected in the recombinant VLPs, quality control experiments were designed to troubleshoot and validate the RdRP incorporation results from MS sequencing by investigating viral RNA expression and characterization of RdRP activities in the following two sections.

MS Protein Identity Sequencing Results for VLPs from Both Viral RNA Segment Expression

Accession	Description	Score	Coverage	# Unique Peptides	# PSMs	# AAs	MW [kDa]	Calc. pI
CP	Capsid Protein (hPBV)	5111.90	86.59	32	1872	552	62.0	6.18
ORF2	ORF2 (hPBV)	930.61	94.20	29	393	224	24.9	8.43
15804574	50S ribosomal protein L1 [Escherichia coli O157:H7 EDL933]	690.39	79.91	19	215	234	24.7	9.64
24053794	50S ribosomal subunit protein L23 [Shigella flexneri 2a str. 301]	278.68	42.00	4	68	100	11.2	9.94
24053796	50S ribosomal subunit protein L3 [Shigella flexneri 2a str. 301]	258.94	68.90	15	106	209	22.2	9.91
15803836	50S ribosomal protein L24 [Escherichia coli O157:H7 EDL933]	234.27	51.92	7	63	104	11.3	10.21
15803846	50S ribosomal protein L4 [Escherichia coli O157:H7 EDL933]	218.41	67.66	12	91	201	22.1	9.73
16131180	50S ribosomal protein L15 [Escherichia coli K12]	192.33	61.81	10	75	144	15.0	11.18
24053779	50S ribosomal subunit protein L6 [Shigella flexneri 2a str. 301]	185.93	61.58	12	56	177	18.9	9.70
16128896	outer membrane porin 1a (Ia;b;F) [Escherichia coli K12]	170.20	72.38	18	66	362	39.3	4.96
15803835	50S ribosomal protein L5 [Escherichia coli O157:H7 EDL933]	166.60	77.09	15	82	179	20.3	9.48
15803842	50S ribosomal protein L22 [Escherichia coli O157:H7 EDL933]	153.72	64.55	9	68	110	12.2	10.23
15803823	30S ribosomal protein S4 [Escherichia coli O157:H7 EDL933]	150.33	67.48	16	66	206	23.5	10.05
24053793	50S ribosomal subunit protein L2 [Shigella flexneri 2a str. 301]	131.20	57.14	15	62	273	29.8	10.93
15799798	pyruvate dehydrogenase (decarboxylase component) [Escherichia coli O157:H7 EDL933]	122.93	37.32	31	49	887	99.6	5.68
24053051	50S ribosomal subunit protein L19 [Shigella flexneri 2a str. 301]	121.92	66.96	8	41	115	13.1	10.62
24051180	30S ribosomal subunit protein S1 [Shigella flexneri 2a str. 301]	112.20	50.63	24	41	557	61.1	4.98
15803840	50S ribosomal protein L16 [Escherichia coli O157:H7 EDL933]	103.27	32.35	3	35	136	15.3	11.22

Table 3.3: Protein identity MS results for recombinant hPBV from the co-expression of both viral RNA segments. The VLPs were yielded from the expression of pETDuet-ΔRBS-hPBV.1. Score: Displays the protein score, which is the sum of the scores of the individual peptides. Coverage: Displays by default the percentage of the protein sequence covered by identified peptides. # Unique Peptides: Displays the number of peptide sequences unique to a protein group. # PSMs: Displays the total number of identified peptide sequences (peptide spectrum matches) for the protein, including those redundantly identified. # AAs: Shows by default the sequence length of the protein. MW [kDa]: Displays the calculated molecular weight of the protein. The Proteome Discoverer application calculates the molecular weight without considering PTMs. Calc. pI: Displays the theoretically calculated isoelectric point, the pH at which a particular molecule carries no net electrical charge.

3.4. Characterize the terminal sequence of expressed viral RNA molecules

Previous studies proposed that RdRP was encapsidated through its interactions with viral RNA segments[83]. Therefore, transcripts from T7 RNA polymerase must contain the exact termini of viral RNA transcripts, as RdRP usually interacts with RNA

termini[83,92]. In addition, packaging signal (PS) sequences are generally located at the terminal untranslated regions of viral RNAs[36,93]. Since T7 RNA polymerase initiates RNA transcription at the juxtaposed G after the TATA sequence in the upstream T7 promoter, RNA transcripts induced by IPTG could only start with GTAAA as the 5'-end of viral RNA. Therefore, it is only necessary to characterize the HDV ribozyme sequence cleavage at the 3'-end of RNA transcripts after transcription termination.

To investigate the self-cleavage activity of the HDV ribozyme sequence at 3'-end, barcoded cDNA for HDV sequences and the viral 3'-termini was synthesized by tagged primers in strand-specific RT-qPCR, considering the *de novo* self-priming of viral RNAs (Fig. 3.4*A*). A barcode sequence of GCTAGCTTCAGCTAGGCATC was added to the 5'-end of the gene-specific reverse primers in reverse transcription. The cDNA synthesized by the tagged primer contains the barcode sequence, while *de novo* self-primed viral cDNA does not. In real-time PCR, one of the detection primers is the barcode sequence, which eliminates the quantification of the cDNA byproduct synthesized by *de novo* priming. Primers for reverse transcription and real-time PCR are listed in Table 2.6.

In the individual expression of hPBV1 and hPBV2, RNA transcripts with the exact viral 3'-termini were detected in a significantly earlier cycle number (Cq) than those with HDV ribozyme sequences (Fig. 3.4*B*). It indicated that the HDV ribozyme sequences were effectively removed after transcription termination. In the co-expression of hPBV1 and hPBV2 from separate plasmids, RNA transcripts with viral 3'-termini were still significantly more than those with an uncleaved HDV ribozyme sequence. Though HDV cleavage efficiency in co-expression was reduced compared to those in individual expression of each RNA segment, hPBV1 and hPBV2 viral transcripts with the exact 3'-ends were still around 18.86 times or 64.74 times more than transcripts with HDV ribozyme sequence ends, which should still be considered as the effective removal of the HDV ribozyme sequence. However, the co-expression of hPBV1 and hPBV2 from separate plasmids produced RdRP-missing recombinant VLPs at 16 ℃ induction, with undetectable RdRP expression level in cell lysate. Co-expression of hPBV1 and hPBV2 from separate plasmids yielded little recombinant VLPs at 37 ℃ induction, implying improper folding of viral proteins at higher induction temperatures. For expression from

the pETDuet vector, the pETDuet-ΔRBS-hPBV1 expressed hPBV1 after the first T7 promoter while the downstream of the second T7 remained unchanged from the empty pETDuet vector. The pETDuet-ΔRBS-hPBV1 expression was a control for the co-expression from pETDuet-ΔRBS-hPBV to examine the self-cleavage activity of the HDV ribozyme sequence after hPBV1, in individual expression versus co-expression. The results showed that the HDV ribozyme sequence after hPBV1 was properly removed when expressing alone from pETDuet-ΔRBS-hPBV1. However, the HDV ribozyme sequence after hPBV1 remained in the co-expression from pETDuet-ΔRBS-hPBV, while the HDV ribozyme sequence at the 3'-end of hPBV2 was effectively cleaved. It illustrated that a long RNA transcript as hPBV1-HDV-T7-hPBV2 was yielded from pETDuet-ΔRBS-hPBV after IPTG induction. The cleavage inefficiency of HDV ribozyme sequence after hPBV1 from pETDuet-ΔRBS-hPBV expression may explain the
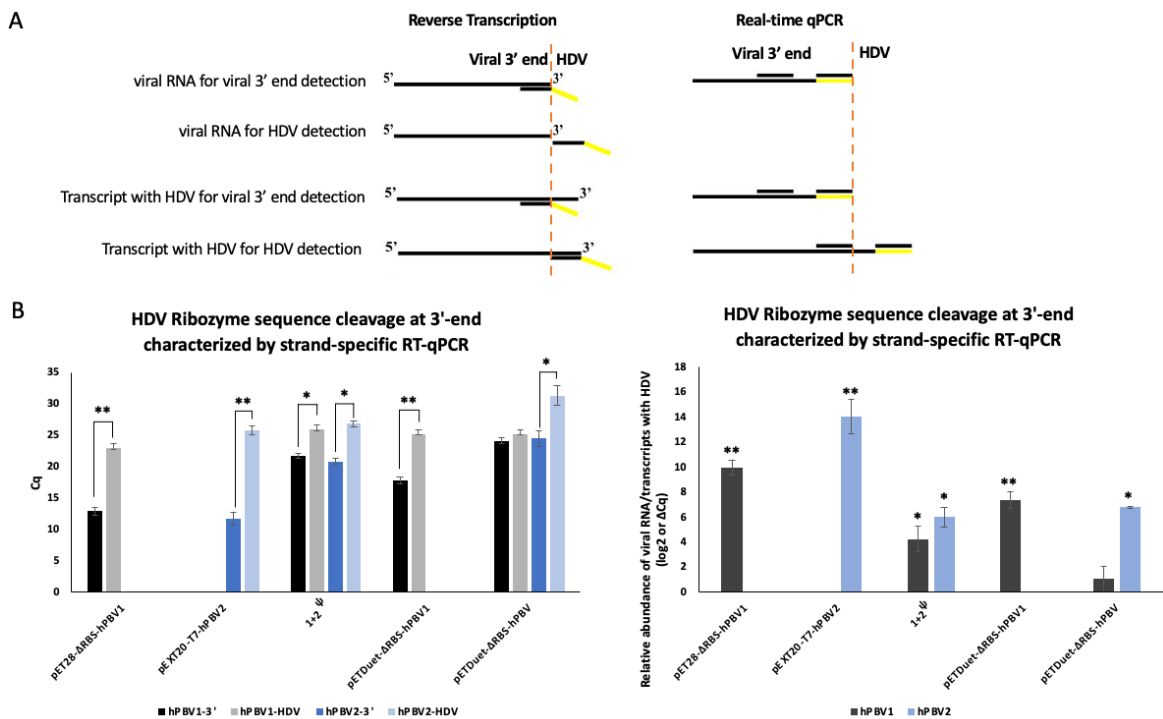
Figure 3.4: Strategy and results for strand-specific RT-qPCR detection of the HDV ribozyme sequence self-cleavage. (A) The scheme demonstration of strand-specific RT-qPCR detection. (B) Strand-specific RT-qPCR results for HDV ribozyme sequence cleavage. The left panel exhibited the raw data of Cq, while the right panel showed the relative abundance of viral RNA over transcripts with uncleaved HDV ribozyme sequences. *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$. ψ: 1+2 sample was the co-expression of pET28-ΔRBS-hPBV1 and pEXT20-T7-hPBV2 at 37 °C.

deficiency of RdRP incorporation in the recombinant hPBV and potentially why the recombinant hPBV was not infectious. Future improvements could optimize the cleavage efficiency of the ribozyme sequences at the 3'-ends of hPBV1 and hPBV2 during co-expression to enhance the yields of both viral RNA transcripts with exact terminates. Alternatively, a T7 terminator sequence could be added to the 3'-end of hPBV1 to promote the HDV ribozyme cleavage during co-expression.

3.5. Characterize the effect of hPBV RdRP activity on the expression of viral RNAs in *E. coli*

Besides the explanation that RdRP incorporation relied on the terminal sequences of RNA transcripts, it could also be possible that RdRP was not detectable due to its low copy number in the assembled VLPs. In addition, the amount of encapsulated RdRP could further decrease depending on the ratio of properly assembled VLPs, as random packaging was a common issue in recombinant VLPs. Therefore, characterizing the RdRP activity is an alternative approach to detect if any RdRP is properly encapsulated into the recombinant VLPs. Assuming a few recombinant VLPs properly encapsulated several copies of RdRP, these RdRPs would synthesize -ssRNA inside the VLP and convert the genome to dsRNA form. Thus, RdRP incorporation could be demonstrated by detecting -ssRNA from strand-specific RT-qPCR, which was not transcribed by T7 RNA polymerase. In addition, after recovering the dsRNA genome, the intact mature virus particles could synthesize more viral mRNA, resulting in more translated viral proteins as an amplified signal of properly packaged RdRP.

Before the characterization of RdRP activities, it was noted that RdRP translated from hPBV2 was reduced during co-expressing with hPBV1. Further investigation showed that the RdRP protein translation from hPBV2 behaved slightly differently between the individual expression and its co-expression with hPBV1. While the induction temperature did not significantly affect RdRP expression levels when hPBV2 was expressed alone (Fig. 3.5.1*A*), RdRP translation preferred higher induction temperatures during co-expression (Fig. 3.5.1*B*). However, poor VLP yield was observed from the

37 °C co-expression, presumably due to improper protein folding. It could be possible that the upstream inherent RBS of RdRP might behave more actively at a higher induction temperature and thus could compete over the upstream RBS of CP on hPBV1 (Fig. 1.2.2.2). Alternatively, *E. coli* may not be the natural host of hPBV, RdRP protein expression regulation by its upstream RBS might behave differently in its native hosts.
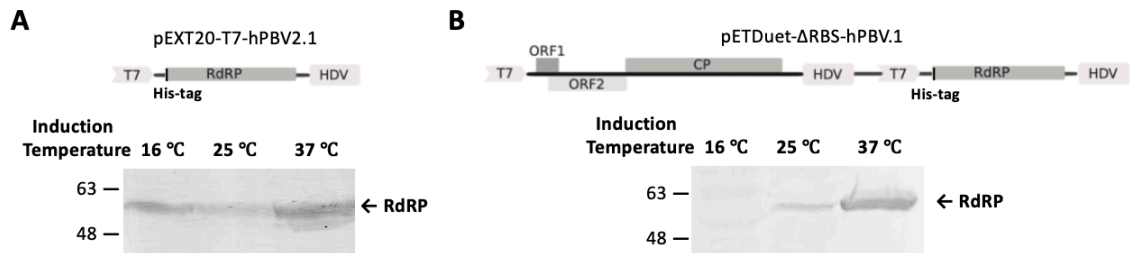
Figure 3.5.1: Western blot detections of the RdRP expression levels at different induction temperatures. (A) RdRP expression from hPBV2 at 16 °C, 25 °C, and 37 °C induction temperatures. The loaded samples were Ni-NTA eluted fractions from 1L of pEXT20-T7-hPBV2.1 expression culture detected by the anti-His monoclonal antibody (Invitrogen). (B) RdRP expression from hPBV2 by pETDuet-ΔRBS-hPBV.1 at 16 °C, 25 °C, and 37 °C induction temperatures. Similarly, the Ni-NTA elution fractions were loaded and detected on the anti-His Western blot.

As a negative control for RdRP activity characterization, the RdRP motif C (-GDD-) was mutated to a dysfunctional mutant (-GAA-) to inactivate the replicase. RdRP motif C was known to play an essential role in rNTP binding during RNA synthesis[94,95].

Viral RdRP activity was first characterized by its effect on viral protein expression, which was considered an amplified signal for properly incorporated RdRP. One properly assembled recombinant virus particle could synthesize many copies of viral mRNA and sequentially translated viral proteins. His-RdRP detected by anti-His Western blots represented relative viral protein levels from hPBV1 and hPBV2 co-expression where hPBV2 encoded either wild-type RdRP or -GDD- to -GAA- RdRP mutant. Western blot results approved that more viral protein expression was detected in the co-expression where hPBV2 encoded wild-type RdRP (Fig. 3.5.2*A-B*). Western blot results were then supported by the corresponding RNA expression levels analyzed by strand-specific RT-qPCR of total RNA in the lysate(Fig. 3.5.2*C*)[96,97], where both +ssRNA and -ssRNA of both hPBV1 and hPBV2 were relatively more abundant in the co-expression lysate where hPBV2 encoded wild type RdRP (Fig. 3.5.2*D*). Consistency of Western blot

and RT-qPCR results implied that higher viral protein yield from the co-expression of hPBV1 and wild-type RdRP encoded hPBV2 was due to the abundance of corresponding RNA transcripts. To be noted, The -ssRNA of hPBV1 and hPBV2 was observed at a highly deficient level (average $Cq > 29$ cycles) in the co-expression where hPBV2 encoded dysfunctional RdRP mutant. The relative abundance of -ssRNA in the co-expression containing wild-type RdRP implied that few recombinant VLPs properly encapsulated viral RdRP, replicated -ssRNA, and actively transcribed more viral mRNA for viral protein expression.
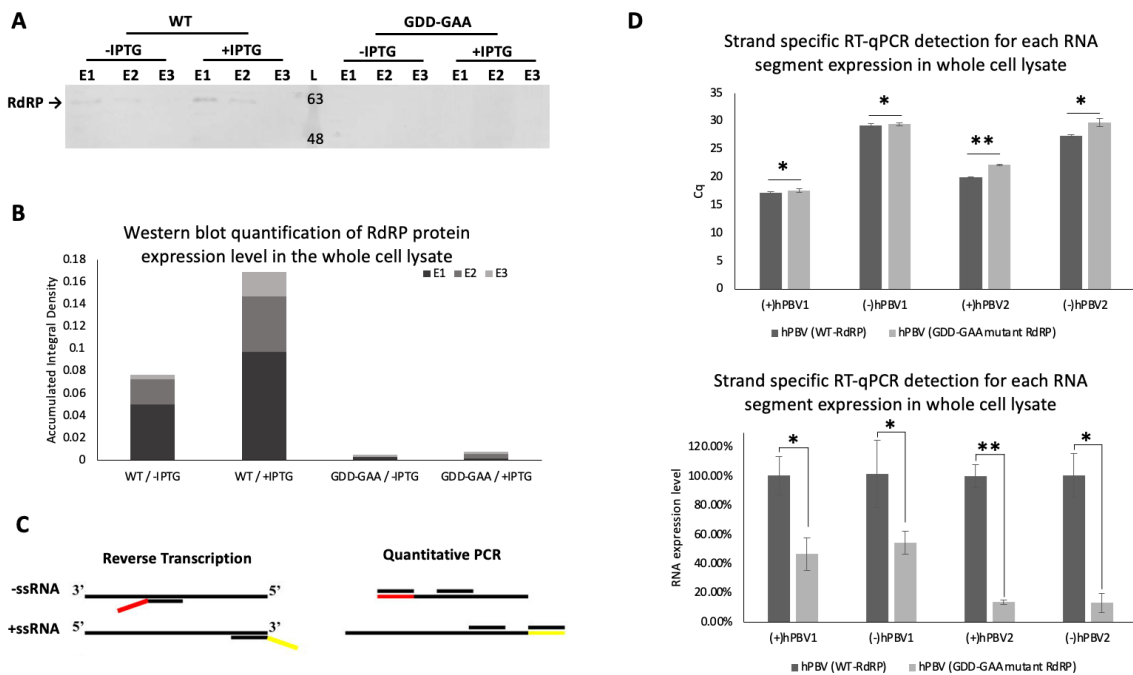
Figure 3.5.2: Western blot quantification and RT-qPCR results to estimate RdRP activities in the hPBV segment expression. (A) Western blot results for His-tagged RdRP detection in the co-expression with wild-type RdRP or mutant RdRP. WT: co-expression of both segments from pETDuet-hPBV.1; GDD-GAA: co-expression with the mutant RdRP (-GDD- to -GAA-) from pETDuet-hPBV.2. (B) Stack plot of the Western blot quantification results. It indicates the accumulative amounts of signals. (C) The strategy of strand-specific RT-qPCR. (D) Strand-specific RT-qPCR results. The upper panel showed the raw Cq results. The lower panel interpreted WT RdRP encoded from hPBV2 as 100% expression level and indicated the relative RNA levels for dysfunctional RdRP mutant encoded from hPBV2. Statistical significance: * for $p < 0.05$ and ** for $p < 0.01$.

3.6. Summary

All predicted viral proteins were successfully detected upon individual viral RNA expression under IPTG induction. Translated viral proteins assembled into recombinant hPBV virus-like particles mainly composed of CP and ORF2. ORF1 had a decent expression level as a soluble protein but was not encapsulated in the recombinant VLPs. RdRP protein translation from hPBV2 was affected and suppressed when co-expressing with hPBV1, potentially due to an unknown translation regulation mechanism of the viral RNA by different upstream inherent RBSs. The translation regulation mechanism usually allows a higher CP protein expression level than RdRP, considering each hPBV particle was proposed to be made of 120 copies of CP but only 1-2 copies of RdRP[83,98,99]. Plasmid-based reverse genetic systems for mammalian reoviruses and rotaviruses, where viral RNA segments were also expressed from their cDNA sequences flanked by a T7 promoter and an HDV ribozyme sequence in cloned plasmids, produced infectious recombinant VLPs from mammalian cell culture with RdRP properly incorporated[100,101]. Therefore, quality control experiments characterized the hPBV expression by the cleavage efficiency of HDV ribozyme sequence at the end of RNA transcripts and the RdRP activities on the viral RNA and protein levels if minor recombinant hPBV were correctly assembled. Overall, HDV ribozyme sequences were effectively removed after transcription, with a potential long RNA transcript of hPBV1-HDV-T7-hPBV2 yielded from hPBV1 and hPBV2 co-expression from a pETDuet vector.

Further optimization could manipulate the ribozyme sequences at the end of RNA transcripts to promote efficient self-removal. In addition, the RdRP activity characterization concluded that few recombinant hPBV virus particles encapsulated enzymatically active RdRP, which replicated the -ssRNA and transcribed more viral mRNA resulting in more viral protein expression. Future experiments could optimize the following parameters: (1) ribozyme sequences for self-cleavage after transcription termination to release the exact termini of viral mRNA; (2) inherent upstream RBS sequences to manipulate the translation regulation of each ORFs; (3) viral RNA expression in the natural host of hPBV after unveiling its native host(s).

<div align="right">

Chapter 4

</div>

Characterization of viral RNA segments in the recombinant VLPs

For many dsRNA viruses, +ssRNA transcripts not only play an essential role as mRNA in viral protein translation but also are required to be encapsulated to replicate RNA genomes in progeny virus particles. In reoviruses, it is proposed that each transcription enzyme complex (TEC) binds to one and only one RNA segment by recognizing the RNA termini and facilitating the packaging of the viral genome into newly assembled particles via genome-TEC-CSP interactions, which is more likely to assembly into an infectious particle[32]. *In situ* cryo-EM structures of different reoviruses with encapsulated RdRP and RNA genome demonstrate the interactions between capsids and RNA genome, RdRP and RNA termini, and capsids and RdRP[14,29,31,32,40]. Furthermore, studies in rotaviruses showed that the RdRP template entry tunnel specifically recognizes the UGUG sequence in the conserved 3' termini of viral RNA for selective viral RNA packaging[102,103]. In bacteriophage φ6, gene segment L encodes viral protein P1 (capsid), P2 (RdRP), P4 (NTPase), and P7 (assembly co-factor), which assemble into procapsids[104]. The +ssRNA segments of S (small), M (medium), and L (large) are sequentially encapsulated in the order of size via the packaging signals located at the 5' untranslated termini[105–107]. The presence of packaging signal (PS) sequences in the viral RNA genome, which are usually found in the 5'- and 3'- untranslated regions and their vicinities, is a ubiquitous phenomenon not only in dsRNA viruses but also in many other RNA viruses. Besides the earlier discussed rotavirus and bacteriophage φ6, the 5'- and 3'-terminal sequences of mammalian *orthoreovirus* (MRV) are essential for

packaging during assembly, as well as reassortment and recombination of RNA genomes during co-infection or mixed infection with other reoviruses[100,108,109]. In yeast L-A viruses, a stem-loop structure at the 3'-end of the +ssRNA is responsible for binding the *pol* domain of the minor coat protein[110]. As for positive-sense single-stranded RNA viruses, a 190-nucleotide sequence in the 5'-end and a 95-nt bulged stem-loop in the 3'-end were identified as the PS sequences in mouse hepatitis virus (MHV)[111–113]. PS sequences of alphaviruses usually include 4-6 stem-loops with a conserved GGG motif at their loops in the viral genome[114]. The number of GGG-containing stem-loops determines the activity of PS function[114]. PS sequences in influenza viruses are highly conserved among strains and located at the 5'- and 3'-termini of every segment, including most untranslated regions on both sides and some adjacent coding regions[93]. As for retroviruses, the ψ region at the 5'-end of HIV-1 viral RNA binds to Gag polyproteins, which assemble into VLPs. In addition, RNA-RNA interactions also determine the packaging order and secure the incorporation of all viral RNA genomes in multi-segmented dsRNA viruses. Previous studies in rotaviruses characterized the intersegment RNA-RNA interactions between the smallest RNA segment 11 and other ten RNA segments, thus proposing a selective packaging model of eleven RNAs which explained the sequence-specific assortment of reovirus RNA genomes[115]. Deletion or mutation of the terminal sequences for intersegment RNA-RNA interactions prevented genomic RNA packaging in blue tongue viruses (BTVs)[116]. Therefore, in this section, I attempted to characterize the incorporation status and the packaging mechanism of the two viral RNA segments in recombinant hPBV VLPs.

4.1. Presence of viral RNA segments in recombinant hPBVs

Previous studies found bacterial rRNA consisted of approximately 95% to 99% encapsulated RNA in virus-like particles assembled by hPBV CPs from recombinant protein expression in *E. coli*[37]. These encapsulated rRNA molecules were proposed to be randomly packaged by the positively charged N-terminus of hPBV CP proteins, which explained why *E. coli* ribosomal proteins were detected in recombinant hPBV as major contaminants[37]. With the recombinant hPBV produced from viral RNA expression, I

attempted to identify the two viral RNA segments in the purified recombinant VLPs. Well-assembled recombinant VLPs secure encapsulated viral RNAs by surrounding capsid proteins, while leakage of genomic RNA potentially activates the host antiviral mechanism, as discussed in section 1.1.1. Since RNA molecules could penetrate into higher CsCl density levels, which potentially co-localized with recombinant VLPs, the encapsulated RNA was isolated by RNA extraction from immunoprecipitated VLP samples against anti-PBV CP polysera. A cDNA library of encapsulated RNA molecules was obtained by reverse transcription against random hexamer primers. To confirm the presence of viral RNA in the library, cDNAs of both viral RNA segments were amplified by gene-specific primers complementary to their 5'- and 3'- terminal sequences. The profile of cDNA amplificants was analyzed by 2% agarose gel electrophoresis. Two bands around 2.5 kb and 1.7 kb were observed, matching the expected sizes of hPBV1 and hPBV2, respectively (Fig. 1.2.3.1 and Fig. 4.1*A*). PCR amplified products of these two bands were sequenced and identified as cDNAs of hPBV1 and hPBV2 by Sanger sequencing (Genewiz).

RT-qPCR analyses provided a high-throughput alternative approach to characterize the presence or absence of each viral RNA segment in the immunoprecipitated VLPs from different expression combinations. As a negative control, no viral RNA segments were found in VLPs from codon-optimized CP expression (pET19b-CP) (Fig. 4.1*B*), consistent with the previous results that most encapsulated RNAs from CP expression were rRNAs[37]. In contrast, hPBV1 was found in VLPs upon hPBV1 expression (pET28a-ΔRBS-hPBV1), indicating CP's capacity to enclose viral RNAs (Fig. 4.1*B*). Both hPBV1 and hPBV2 were detected as encapsulated RNAs in recombinant VLPs after co-expressing the two segments by a pETDuet vector (pETDuet-ΔRBS-hPBV) or from separate plasmids (pET28a-ΔRBS-hPBV1 & pEXT20-T7-hPBV2) (Fig. 4.1*B*). The results suggested that both hPBV1 and hPBV2 were capable of being packaged inside recombinant VLPs. Though whether both two viral RNA segments or only one of them were packaged in each recombinant hPBV particle was not conclusive from the RT-qPCR data, the abundance of viral RNAs detected in immunoprecipitated VLPs at least confirmed that they were encapsulated in the well-assembled recombinant VLPs. Since RdRP was not identified in recombinant hPBVs yielded from hPBV1 and

hPBV2 co-expression (Table 3.3), the detection of encapsulated hPBV1 and hPBV2 suggested that hPBV viral RNA incorporation mechanism was independent of RdRP, presumably directly interacting with the positively charged N-terminal region of CP. The phenomenon was further approved by detecting the packaged hPBV1 in VLPs from pET28a-ΔRBS-hPBV1, where RdRP or hPBV2 was absent. In conclusion, recombinant hPBVs from viral RNA segment expression encapsulated the expressed viral RNA segment(s), the packaging of which was independent of RdRP incorporation.
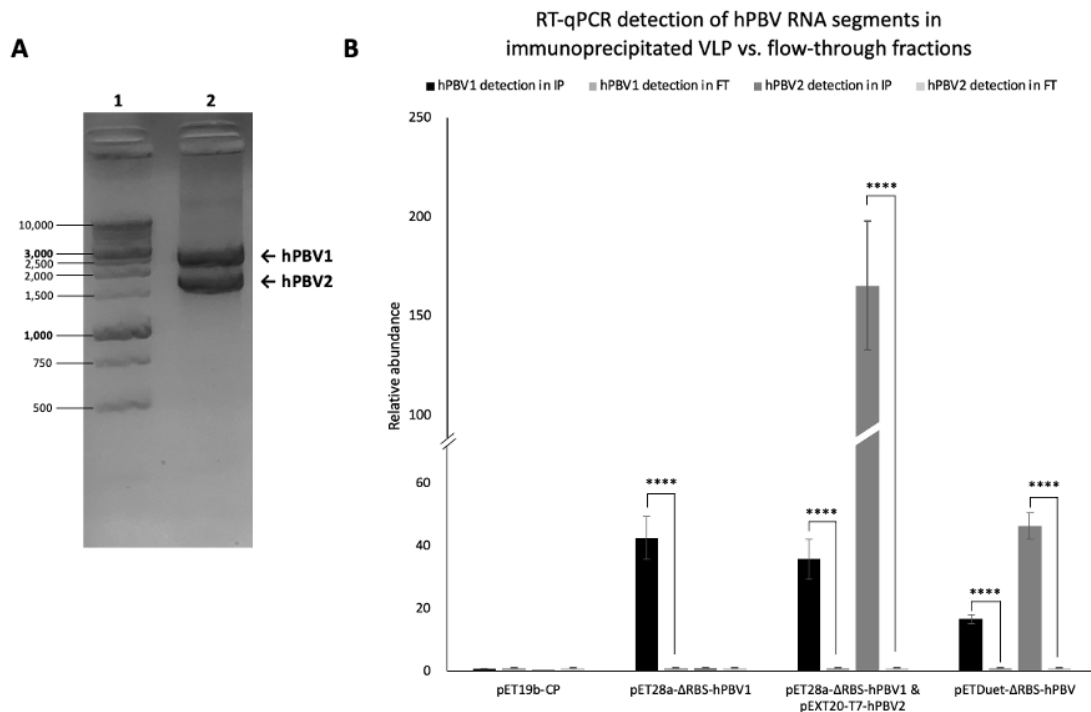
Figure 4.1: Analysis of RNA molecules packaged inside the recombinant VLPs.
(A) The cDNA amplification profile of the RNA extracts from the immunoprecipitated VLPs from co-expression of hPBV and hPBV2. Lane 1: dsDNA ladder (GoldBio D010). Lane 2: amplified cDNA from the VLP sample. The hPBV1 RNA segment is expected to be 2525 bp, and 1745 bp for hPBV2. (B) The RT-qPCR results for immunoprecipitation assay of purified VLPs. RNA extracts of the immunoprecipitated VLPs and corresponding flow-through fractions. The relative abundance was calculated by normalizing the viral RNA detection in the corresponding flowthrough fractions as one arbitrary unit. Statistical significance ****: $p < 0.0001$.

4.2. The packaging mechanism of viral RNA segments

Given that viral RNA segments were observed in recombinant hPBVs, the incorporation of viral RNA segments was characterized to unveil whether viral RNAs were specifically packaged and how they were encapsulated. Since expression of recombinant CP protein resulted in rRNA random incorporation in VLPs, the packaging specificity of viral RNA was compared with that of a host mRNA (*ihfB*), which presumably represented the mispackaged RNAs from random incorporation[117]. The host internal reference gene *ihfB* encodes one of the two subunits of the integration host factor, which regulates genetic recombination, transcription, and translation. Selective packaging efficiency was measured by comparing the relative abundance of viral RNA and ihfB in whole lysate and that in purified immunoprecipitated VLPs. To avoid the biased detection of viral RNA due to *de novo* priming in reverse transcription, viral RNAs were detected by strand-specific RT-qPCR. The expression of codon-optimized CP served again as a negative control, where no viral RNA expression or incorporation
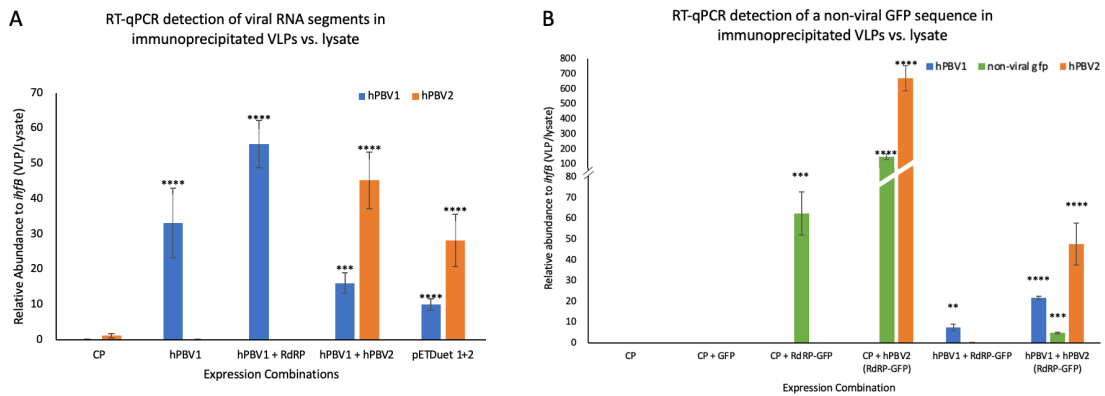
Figure 4.2: Viral RNA selective incorporation mechanism indicated by strand-specific RT-qPCR results. (A) RT-qPCR detection of viral RNA selective packaging. The amount of each detected viral RNA segment was normalized by the internal reference *ihfB*. The relative abundance of viral RNA and *ihfB* were compared between encapsulated RNAs in VLPs and expressed RNAs in lysates. CP was expressed from pET19b-CP, hPBV1 from pET28-ΔRBS-hPBV1, and hPBV2 from pEXT20-T7-hPBV2. The pETDuet 1+2 represents the co-expression from pETDuet-ΔRBS-hPBV. (B) RT-qPCR detection of the selective packaging of a non-viral GFP sequence. The relative abundance was normalized by *ihfB* as in (A). GFP was expressed from pET19b-GFP, RdRP-GFP from pET19b-GFP, and hPBV2 (RdRP-GFP) from pEXT20-T7-hPBV2.3. Statistical significance can be referred as following: *****: p < 0.05, ******: p < 0.01, *******: p < 0.001, ********: p < 0.0001.

was detected. In the expression of hPBV1, the hPBV1:*ihfB* ratio was increased by 33.18 times in immunoprecipitated VLPs from total expressed RNAs in lysate, while hPBV2 was not detected at all due to no hPBV2 expression (Fig. 4.2*A*). Since only hPBV1 was expressed, it indicated that hPBV1 was sufficient to be selectively packaged alone without the presence of RdRP or hPBV2. It further implied that the selective incorporation of hPBV1 was merely based on the RNA-protein interactions between viral RNA and CP. In the co-expression of hPBV1 and codon-optimized RdRP, the hPBV1:*ihfB* ratio was enhanced by 55.46 times from the expressed RNAs in lysate to the encapsulated RNAs in recombinant VLPs (Fig. 4.2*A*). As expected, no hPBV2 was detected as the detection primers annealed to the 3'-untranslated termini, which was removed in the codon-optimized RdRP protein expression. Compared with hPBV1 expression alone, an increased hPBV1:*ihfB* ratio in hPBV1 and RdRP co-expression was observed, implying a higher percentage of recombinant VLPs encapsulated viral RNAs in the presence of RdRP. It remained for future experiments to characterize how RdRP facilitated the incorporation of viral RNA in recombinant VLPs. In the co-expression of two viral RNA segments from a single pETDuet vector or two separate plasmids, hPBV1 and hPBV2 were both detected with significantly increased abundances in recombinant VLPs as capsulated RNAs compared with randomly packaged host *ihfB* mRNA (Fig. 4.2*A*). However, referring to similar expression levels of hPBV1 and hPBV2 in the whole lysate from co-expression (Fig. 3.4*B*), hPBV2 was more abundant than hPBV1 in the encapsulated RNAs of recombinant VLPs from co-expression (Fig. 4.1*B* and Fig. 4.2*A*). Due to the limitation of characterizing encapsidated RNA components at the single particle level, the explanation could be either some recombinant particles packaged extra copies of hPBV2 or some VLPs only encapsulated hPBV2. To sum up, the results concluded that viral RNA segments were selectively packaged into recombinant hPBV VLPs.

From previous studies, selective packaging of viral RNA relies on conserved RNA sequences that interact with CP and RdRP, which are usually characterized as packaging signal (PS) sequences[100–102,106,118]. To investigate whether encapsulated viral RNAs in recombinant hPBVs were selectively packaged via PS sequences, I monitored the packaging of a non-viral GFP sequence (*gfp*) as an independent sequence or an

insertion in hPBV2. Since PS sequences were usually located at the 5'- and 3'-untranslated regions in viral RNA termini[36], the non-viral GFP sequence was inserted at the end of the RdRP coding region. Successful expression of *gfp* was marked by green fluorescence under UV light due to the translated fluorescent GFP protein. The expression of codon-optimized CP served as the negative control (Fig. 4.2*B*). No encapsulation of *gfp* or viral RNA was detected since there was no corresponding RNA expression. Surprisingly, in the co-expression of codon-optimized CP and *gfp*, the *gfp* sequence was less preferred to be packaged than host mRNA *ihfB*, with the *gfp*:*ihfB* ratio of relative RNA abundance in VLPs versus lysate as 0.047 (Fig. 4.2*B*). If an RNA molecule was packaged at the same efficiency as *ihfB*, its relative RNA abundance in VLPs versus lysate would be the same as that of *ihfB*, i.e., ratio$_{rna:ihfB}$ = 1.0. A *gfp*:*ihfB* ratio of 0.047 showed that the non-viral *gfp* was unlikely to be incorporated in recombinant VLPs as an independent sequence. When co-expressing codon-optimized CP and RdRP-GFP, *gfp* was significantly more preferred to be encapsulated than *ihfB* (Fig. 4.2*B*), indicating that *gfp* insertion in an RNA transcript encoding hPBV RdRP interacted with CP at a higher affinity than host mRNA. However, in the co-expression of hPBV1 and RdRP-GFP, a *gfp*:*ihfB* ratio of 0.282 implied that the RdRP-encoded RNA could no longer carry the non-viral *gfp* into VLPs in the presence of wild-type viral RNA, i.e., hPBV1 (Fig. 4.2*B*). Compared the results from CP and RdRP-GFP co-expression and those from hPBV1 and RdRP-GFP co-expression, wild-type viral RNA exhibited a higher interaction affinity to capsids than a codon-optimized RNA transcript (RdRP-GFP), suggesting that PS sequences were present in the native viral RNA sequences. Sequence modifications, including removing the 5'- and 3'-untranslated regions and codon-optimizing the coding region, disrupted the function of PS sequences. Interestingly, *gfp* insertion in hPBV2 displayed preferential incorporation of the non-viral *gfp*, with a significantly higher relative abundance of *gfp* and hPBV2 in VLPs versus lysate than that of *ihfB* from the co-expression of CP and *gfp*-inserted hPBV2 (Fig. 4.2*B*). The selective incorporation of hPBV2 in CP and hPBV2 co-expression demonstrated that hPBV2 was able to be selectively packaged in the absence of hPBV1. Combined with previous results that hPBV1 was selectively encapsulated in recombinant VLPs from hPBV1 and RdRP co-expression (Fig. 4.2*A*), hPBV was able to encapsidate hPBV1 and

hPBV2 separately and selectively. The selective packaging of *gfp* inserted in hPBV2 was still observed in the co-expression of hPBV1 and *gfp*-inserted hPBV2, illustrating that the presence of PS sequences on hPBV2 was able to carry a non-viral sequence into recombinant VLPs as efficient as native viral RNAs. Compared the *gfp* selective packaging results in a *gfp*-inserted hPBV2 with those in a codon-optimized RdRP-GFP transcript, it concluded that the selective packaging of viral RNA relied on the presence of PS sequences. Deletion and mutation of PS sequences could cause the RNA transcript to be less competitive in the interactions with capsids.

4.3. Summary

In recombinant hPBV particles, viral RNA segments were selectively packaged if they were present in the host cells. In addition, the two viral RNA segments could be separately incorporated into recombinant VLPs, and hPBV1 could be packaged in dependent of RdRP. The presence of intact packaging signal (PS) sequences was required for selective packaging at maximal efficiency. Since the presence of PS sequences on hPBV2 granted the selective packaging of a non-viral sequence into VLPs, I proposed that hPBV2 incorporation merely relied on RNA sequences independent of RdRP expression. Further experiments could characterize hPBV2 incorporation by expression hPBV2 with a deletion of internal sequences that encoded functional RdRP domains. Since the template entry tunnel of RdRP was responsible for RNA packaging in rotaviruses[103], mutations of residues interacting with RNA molecules in the hPBV RdRP template entry tunnel could be an alternative approach to characterize if RdRP facilitated the encapsulation of hPBV2. In addition, due to the limitation of analyzing the encapsulated RNA sequence at the single virus particle level, two scientific questions of detailed mechanism remained unknown: (1) if the two RNA segments were packaged in an equimolar ratio, or one of the two was incorporated with a higher copy number; (2) if any recombinant VLPs only incorporated one of the two RNA segments since hPBV1 and hPBV2 were able to be packaged into capsids separately when expressing alone. Further experiments of high-throughput RNA sequencing would demonstrate the percentage of recombinant hPBV particles containing full-length viral RNA. The

sequences of PS signals on the two viral RNA segments could be further investigated by manipulating hPBV1 and hPBV2 expression sequences in future studies.

<div align="right">

Chapter 5

</div>

Characterization of ORF2 in the recombinant VLPs

ORF1 and ORF2 encoded by hPBV1 are the two functionally uncharacterized viral proteins in the hPBV genome. According to the results in Chapter 3, both ORF1 and ORF2 viral proteins were successfully translated from the upstream inherent RBS sequences in *E. coli* Rosetta 2 cells. Still, only ORF2 was found in the yielded recombinant VLPs. In previous studies, ORF2 was once observed as a 24.9 kDa protein from a cell-free *in vitro* transcription and translation system, but CP expression was not detectable in the same experiment[82]. Though the function of ORF2 has not yet been proposed, a matrix-based sequence comparison shows that three phylogenetically distant PBVs, two human PBVs and one rabbit PBV, harbors conserved ExxRxNxxxE motifs repeated 4-10 times, which were predicted to be mainly unfolded[119]. On the other hand, the C-terminal domain of ORF2 is hydrophobic with a glycine-rich region of 25-40 residues[119]. Since ORF2 was observed in the purified recombinant hPBVs, investigating the ORF2 packaging mechanism could illustrate the function of ORF2 in the hPBV life cycle. This chapter characterized the expression and solubility of ORF2 protein and its sequential truncated mutants. The presence or absence of different truncated ORF2 proteins in recombinant VLPs mapped out the interaction region of ORF2 for ORF2-CP interactions, which were responsible for ORF2 incorporation.

5.1. The incorporation status of ORF2

According to the MS protein identity sequencing results of different VLPs (Table 3.3), ORF2 protein was found in the purified recombinant VLPs. The procapsid assembly of bacteriophage φ6 was characterized *in vitro* by mixing the individually purified procapsid components (P1, P2, P4, and P7) at different molar ratios[104]. In an attempt to generate an *in vitro* assembly system for recombinant hPBV, ORF2 protein was expressed alone from pET-SpecR-ORF2-FL (Table 2.1.3). Unexpectedly, the ORF2 protein was insoluble, and the insoluble ORF2 in the pellet was unable to be solubilized by mixing with purified CP proteins, indicating that ORF2 was incorporated along with the synthesis and folding of CP. To eliminate the possibility that ORF2 might precipitate due to its binding to an insoluble host protein, His-tagged ORF2 protein was purified by Ni-NTA resins in buffers with 8 M urea and eluted after the on-column refolding, which gradually decreased urea concentration in wash buffers to 1 M. His-tagged ORF2 was eluted with high purity from Ni-NTA column but quickly aggregated into white precipitants after elution (Fig. 5.1.1). The purification results implied that ORF2 protein was insoluble due to its poorly folded structure in solution. As an alternative approach to characterize insoluble proteins with disordered domains, a stable and soluble truncated mutant was usually generated for functional studies based on the computationally predicted structure of the full-length protein. However, as shown in previous analyses,
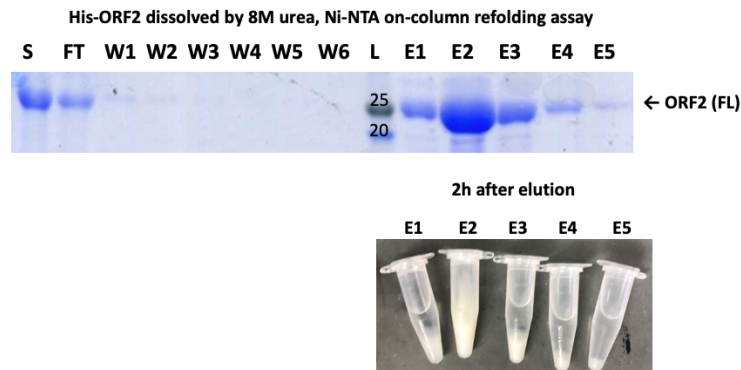

Figure 5.1.1: ORF2 purification through 8 M urea denaturation and on-column refolding. S: Soluble fraction in 8 M urea; FT: flowthrough; W1-W6: washes in buffers with decreasing urea concentrations to 1 M. E1-E5: 1 ml elution fractions in buffers containing 250 mM imidazole and 1 M urea.

the predicted structures of ORF2 by different algorithms mainly remained unfolded[119].
Despite a single long α-helix predicted by AlphaFold2, the AWSEM algorithm predicted
a major long α-helix of the N-terminal domain with a short turn of α-helix at its C-
terminus (Table 5.1)[120,121]. The prediction of AWSEM interpreted the structural features
in previous amino-acid sequence analyses: an unfolded N-terminal domain with repeated
ExxRxNxxE motifs and a hydrophobic C-terminal domain. Summarizing the results from
different structural prediction algorithms, ORF2 was divided into the N- and C-terminal
domains, but no informative prediction in the ORF2 tertiary structure. Therefore, two
sequential truncated mutants were generated from the N- and C-termini, respectively,
with a space of around 60 residues in a total of 244 residues. Two N-terminal truncated
mutants were ORF2[38-224] and ORF2[86-224], and two C-terminal truncated mutants
as ORF2[1-107] and ORF2[1-163] (Table 5.1). To easily detect them in the co-expression
with hPBV1 or CP, the N- and C-terminal truncated ORF2 proteins were fused with an
N-terminal His tag. Each truncated mutant was first expressed alone for solubility tests.
Unfortunately, none of them was soluble (Table 5.1). The insolubility of all truncated
mutants was anticipated from the structural prediction of ORF2 protein, which was
merely an unfolded long α-helix. To confirm the function of ORF was not altered with an
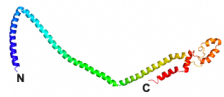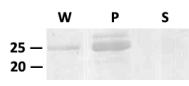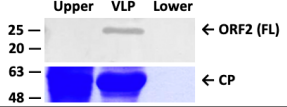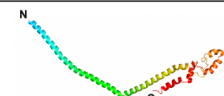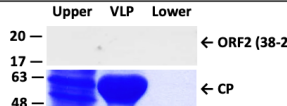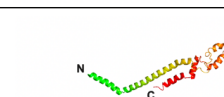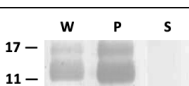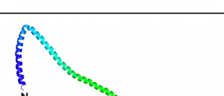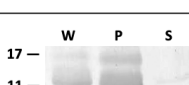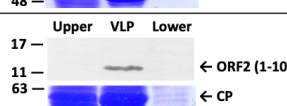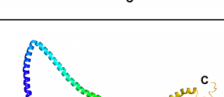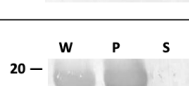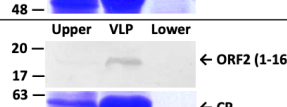
Table 5.1: Expression and solubility summaries of ORF2 and all truncated ORF2.
N: N-termini; C: C-termini; W: Whole lysate; P: Pellet; S: Soluble fraction; Upper: 1
ml fraction located above the VLP fraction; VLP: VLP fraction after CsCl gradient
ultracentrifugation; Lower: 1 ml fraction located below the VLP fraction.

additional N-terminal His tag, i.e., the incorporation of ORF2 while CP synthesis and folding, His-tagged ORF2 was co-expressed with wild-type hPBV1, which encoded a wild-type ORF2. After VLP purification from CsCl gradient ultracentrifugation, the His-tagged ORF2 was only observed in the VLP fraction, without diffusion into adjacent fractions where assembled VLPs were not found (Table 5.1). Co-elution of His-tagged ORF2 proteins in recombinant VLPs showed that a His-tagged ORF2 was encapsidated as efficiently as a wild-type ORF2. Thus, the N-terminal His tag did not interrupt ORF2 incorporation in VLPs, i.e., ORF2-CP interactions.

Inspired by the *in situ* cryo-EM structures of packaged viral RdRP in many dsRNA viruses, I attempted to resolve the *in situ* ORF2 structure inside the recombinant VLPs by cryo-EM. Recombinant VLPs were produced from the co-expression of hPBV1 and His-tagged ORF2, which saturated the incorporation of ORF2 proteins in each recombinant particle by ORF2 overexpression. Occupancies of ORF2 at all possible binding sites in recombinant VLPs were theoretically favored for asymmetric local structure reconstruction in cryo-EM, as more subparticle images were available for different classes of orientations. After subtracting CP signals for subparticle extraction from the collected cryo-EM images, a spherical layer of proteins was observed in diameter of ~ 272 Å (27.2 nm), indicating packaged ORF2 proteins were organized at the
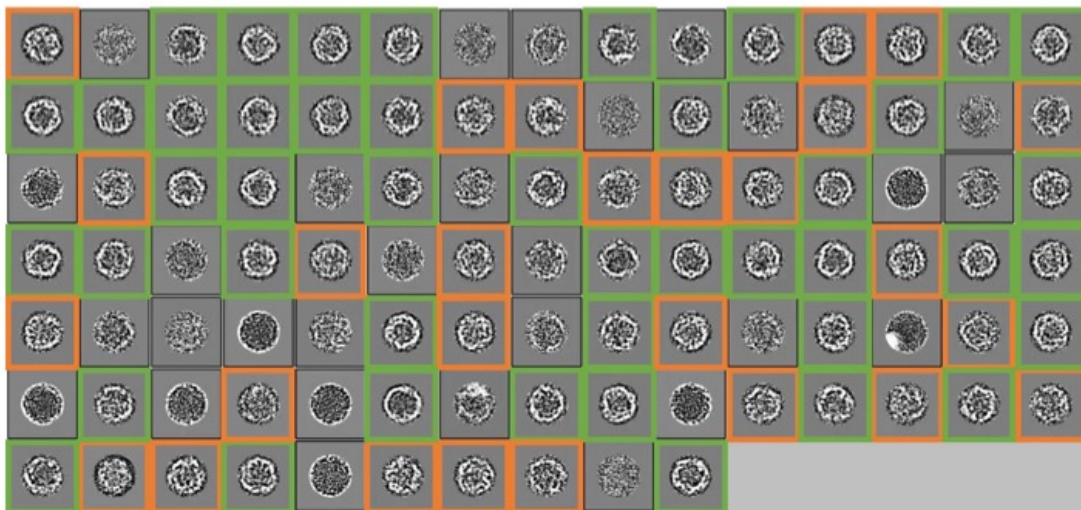
Figure 5.1.2: CryoEM images of ORF2 containing VLPs after CP subtraction. Images marked with green rectangles showed a spherical protein layer underneath the CP was observed. Images marked in orange rectangles showed particles filled with protein densities underneath CP.

interior side of virus capsids (Fig. 5.1.2). Some of the subparticles were filled with protein densities, presumably randomly packaged ribosomal proteins, which were the primary contaminants in recombinant VLPs (Table 3.3). However, the resolution of the ORF2 structure was limited due to the disordered structures of ORF2 in subparticles. As an insoluble disordered protein, it was also tricky to structurally characterize ORF2 or its mutant by X-ray crystallography. Therefore, the function of ORF2 was characterized by biochemical assays for protein-protein interactions, unveiling its incorporation mechanism to virus particles.

5.2. The incorporation mechanism of ORF2

With a plasmid-based expression system for recombinant viruses, it is easy to characterize the function of an interested viral protein by genetically manipulating the viral protein. For example, a deletion of the C-terminal region of rotavirus NSP1 demonstrated the role of NSP1 in promoting the degradation of host IFN regulatory factor 3 (IRF3) and thus antagonizing the innate immune response[101]. To investigate the incorporation mechanism of ORF2 protein, truncated ORF2 mutants, i.e., deletions in N- and C-terminal regions, were co-expressed with wild-type hPBV1 and compared their behaviors with the wild-type ORF2 protein encoded by hPBV1 in the same expression experiment. All truncated ORF2 proteins were fused with N-terminal His tags, which were distinguishable by different molecular weights on anti-His Western blots: His-ORF2[38-224] was at around 21.63 kDa, His-ORF2[86-224] at ~ 15.79 kDa, His-ORF2[1-107] at ~ 13.44 kDa, and His-ORF2[1-163] at ~ 19.54 kDa. As characterization for encapsulated RNA segments in Chapter 4, the purified VLPs from CsCl density gradient ultracentrifugation were immunoprecipitated against anti-PBV CP polysera for ORF2 incorporation characterization in well-assembled recombinant VLPs. The immunoprecipitated VLPs containing different ORF2 truncated mutants were first analyzed on SDS-PAGE for the total protein profiles, ensuring the antibody specificity in immunoprecipitation assays. The presence or absence of an ORF2 truncated mutant was examined by anti-His Western blots and interpreted by signal quantification via ImageJ.

As results showed, co-expression of hPBV1 and full-length His-ORF2 was a positive control. Under an overexpression of ORF2 protein, incorporated full-length ORF2, including the wild type and the His-tagged, was quantified as 17.2% of the total CP protein amount on SDS-PAGE (Fig. 5.2.1*B*). The average amount of incorporated wild-type ORF2 was around 7.6% to 9.3% of total capsid proteins. Given that each virus particle was assembled by 120 copies of CP, the average copy number of wild ORF2 protein was around 9.12 to 11.16, and the saturated copy number of ORF2 was around 20.64. Based on the copy number of incorporated ORF2 and the icosahedral organization of viral capsids, ORF2 was likely distributed at the twelve 5-fold vertices, filling the 5-fold vertices as dimers at its overexpression. Alternatively, ORF2 was distributed at the twenty 3-fold vertices and was half occupied in typical cases. A higher relative ratio of full-length incorporated ORF2 was detected by His-tagged ORF2 on Western blots, presumably due to the signal amplification of primary and secondary antibodies in
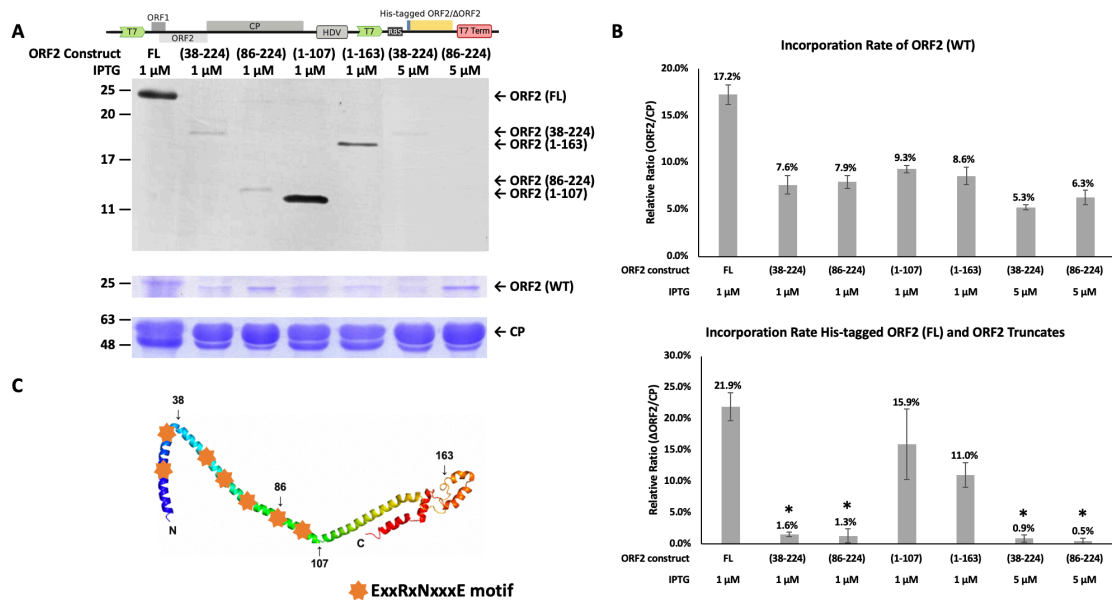
Figure 5.2.1: Western blot quantification for the incorporation rates of ORF2 and truncated ORF2 proteins in the recombinant hPBV VLP immunoprecipitants. (A) The immunoprecipitation (IP) fractions from the purified VLPs by different hPBV1-ORF2 co-expression. The top panel shows the anti-His Western blot for the tagged ORF2 and mutants. The middle panel shows wild-type ORF2 from the hPBV1 expression. The bottom panel shows loaded VLP amounts indicated by the thickness of CP bands. (B) Quantitative analysis of SDS-PAGE for the wild-type ORF2 and the tagged truncated ORF2 incorporation rates into capsids. (C) The computational structure prediction of ORF2 generated by AWSEM-suite. The arrows indicate the position of each truncate. * represents statistical significance with $p < 0.01$.

Western blots. Both C-terminal truncated mutants, His-ORF2[1-107] and His-ORF2[1-163], were incorporated in recombinant VLPs as efficiently as a wild-type ORF2 protein, which was consistently observed in the purified VLP samples from CsCl density gradient ultracentrifugation and the immunoprecipitated VLPs (Table 5.1 and Fig. 5.2.1). The consistency indicated that most co-eluted ORF2 proteins were indeed packaged in the recombinant VLPs. However, the incorporation of N-terminal truncated mutants, His-ORF2[38-224] and His-ORF2[86-224], were significantly disrupted, while the incorporation of the wild-type ORF2 in the same batch was not generally affected (Fig. 5.2.1*B*). Rescue for the deficient encapsulation of the N-terminal truncated mutants was attempted by inducing the protein expression with a higher IPTG concentration (from 1 μM to 5 μM), which resulted in a higher ORF2 expression level[122]. Unfortunately, the incorporation of N-terminal truncated ORF2 was not recovered by ORF2 overexpression, implying that the presence of ORF2 in recombinant VLPs was independent of the available concentrations during assembly but determined by the protein-protein interactions between CP and ORF2. Therefore, the ORF2 N-terminal domain was responsible for CP-ORF2 interactions and determined ORF2 incorporation ability.

Figure 5.2.2: Secondary structure prediction of ORF2. Each line is organized as residue number in the first row, the amino acid sequence in the second row, and secondary structure annotation in the last row. In secondary structure annotation, H stands for α-helix and E for β-strand. The ExxRxNxxxE motifs are highlighted in orange. Arrows indicated truncation sites.

Further analysis of the amino acid sequence of ORF2 N-terminal domain in alignment with the two N-terminal truncated mutants, ORF2[38-224] deleted the first two of the seven repeated ExxRxNxxE motifs, and ORF2[86-224] discarded six of them (Fig. 5.2.1*C* and Fig. 5.2.2). Since the seven ExxRxNxxE motifs were enriched in the N-terminal 102 residues, His-ORF2[107-224] (~ 13.34 kDa) was cloned and proposed to abandon its interaction and incorporation to hPBV capsids completely. Immunoprecipitated VLPs from the co-expression of codon-optimized CP and His-ORF[107-224] were characterized by anti-His Western blots, with a positive control from the co-expression of codon-optimized CP and full-length His-ORFs. Full-length His-ORF2 was co-immunoprecipitated with CP, indicating the incorporation of ORF2 only relied on protein-protein interactions between CP and ORF2. As expected, no truncated ORF2 was observed in VLPs from the co-expression of codon-optimized CP and His-ORF2[107-224] and in the final immunoprecipitated sample.
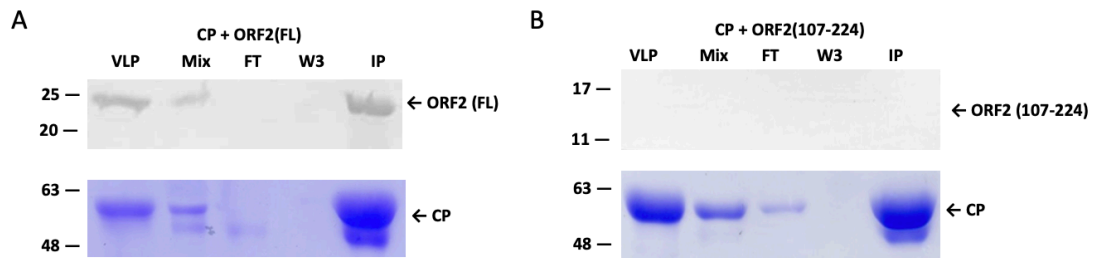
Figure 5.2.3: SDS-PAGE and Western blots for immunoprecipitated VLPs from the co-expression of CP and ORF2. In (A) and (B), the upper panels were anti-His Western blots, and the lower panels were corresponding SDS-PAGE gel stained the Coomassie blue. VLP: VLPs from CsCl density gradient ultracentrifugation; Mix: VLPs mixed with protein A/G magnetic beads and anti-PBV CP polysera; FT: flowthrough; W3: third wash-through; IP: immunoprecipitant.

5.3. Summary

ORF2 was characterized as an unfolded protein that was incorporated into hPBV capsids via the interaction with CP during the translation and folding of CP. The disordered ORF2 can be divided into an N-terminal domain containing seven conserved ExxRxNxxE repeated motifs and a hydrophobic C-terminal domain. The conserved ExxRxNxxE motifs at the N-terminal domain were responsible for the CP-ORF2

interactions and the encapsulation of ORF2 proteins. Disrupting all ExxRxNxxE motifs resulted in the complete removal of ORF2 protein in the assembled recombinant hPBV particles. As a disordered protein, the structure of ORF2 was not resolved due to the resolution limitation in cryo-EM imaging or its insoluble property in X-ray crystallography. Since ribosomal proteins were major contaminants in recombinant hPBV particles, establishing a system to sort and purify homologous ORF2-incorporated recombinant VLPs would enhance the imaging quality of cryo-EM and thus improve the local resolution for subparticle reconstruction. The structure-property of ORF2 protein resembles intrinsically disordered proteins (IDPs), which lack stable secondary or tertiary structure[123]. IDPs are highly prevalent in eukaryotic cells and potentially associated with diseases. For instance, IDPs were found in many cancer-associated proteins like p53 and some cancer-causing viruses, such as the E6 protein in human papillomavirus (HPV)[124]. Analysis of ORF2 by an IDP prediction algorithm (fIDPnn) showed that ORF2 has an N-terminal domain for protein-protein interactions and a small C-terminal domain for protein-RNA interactions[125]. The N-terminal domain of ORF2 was determined for CP-ORF2 interactions based on the incorporation abilities of different N- and C-terminal truncated ORF2 proteins in capsids, which was consistent with the fIDPnn prediction. Future experiments could investigate the predicted RNA-interacting C-terminal domain. Whether ORF2 facilitated RNA incorporation is still needed to be determined in future studies, according to the results of the RNA packaging mechanism discussed in Chapter 4 and the results of the ORF2 incorporation mechanism characterized in this chapter.

The packaged ORF2 protein was organized underneath the interior surface of capsid proteins, marked as a spherical layer of protein densities with a diameter of ~ 272 Å under cryo-EM. Based on the relative amounts of ORF2 and CP in the immunoprecipitated VLPs quantified by ImageJ, each recombinant hPBV particle contains a mean of around 9.12 to 11.16 copies of ORF2, with a maximum copy number of 20.64. Since ORF2 formed an internal protein layer, ORF2 proteins were proposed to distribute around either the twelve 5-fold vertices or the twenty 3-fold vertices.

Chapter 6

Characterization of RdRP in the recombinant VLPs

Properly packaged, enzymatically active RdRP is necessarily required for infectious virus particle expression. In reoviruses, RdRPs in dsRNA viruses are usually anchored at five-fold vertices by adjacent capsid proteins, demonstrated by *in situ* cryo-EM structures[14,29,31,32,126–128]. Even though a migration model was proposed for bacteriophage φ6 RdRP from 3-fold vertices in unexpanded procapsids to 5-fold vertices in mature capsids, hydrophobic interactions and salt bridges between RdRP and capsid proteins were illustrated from *in situ* cryo-EM structure[39,45,106]. By contrast, no direct protein-protein interaction between RdRP and CP was observed in the co-expression of codon-optimized RdRP and CP[83]. RdRP was proposed to be packaged via RNA-mediated interactions, i.e., RdRP-RNA-CP interactions[83]. Compared with the RdRP transcription and replication model based on CP-RdRP interactions, the immediate question is how hPBV regulates RdRP replication and transcription activities upon external environmental signals (NTP and SAM) in the absence of CP-RdRP interactions[32]. In addition, the hPBV RdRP incorporation mechanism is potentially novel due to the lack of CP-RdRP interactions, e.g., whether RdRPs were packaged at a fixed position and how RdRPs were anchored and orientated. To answer these questions, it is required to establish a stable expression platform that produces RdRP-incorporated recombinant VLPs. Therefore, the encapsidation mechanism of RdRP was characterized in this chapter to facilitate the generation of recombinant VLPs with consistent copies of properly packaged RdRPs.

6.1. The incorporation status of RdRP

In previous studies of CP-RdRP interactions, after co-expression codon-optimized CP and RdRP, the unpackaged RdRP was purified by Ni-NTA column, while the assembled VLPs were purified by ultracentrifugation[83]. However, no co-elution of RdRP and CP was obtained from either purified protein samples[83]. Thus, it was concluded that RdRP did not directly interact with CP but presumably via viral genomic RNA[83]. However, RdRP was still missing in the purified VLPs from the co-expression of the two viral RNA segments (Table 3.3). Compared with the previous CP-RdRP interaction studies, the expression level of RdRP from hPBV was highly deficient than the expression as codon-optimized recombinant protein (Fig. 3.5.1). In addition, as discussed earlier in Chapter 3, the packaged RdRP may fall under the detection limitation, given that a proposed CP:RdRP ratio per particle was 120:1 to 60:1 and the presence of empty or random-packaged particles was a common problem in recombinant VLPs. Therefore, I developed a solution to reinforce the hPBV RdRP expression by a separate plasmid that encoded a codon-optimized RdRP. The supplemented RdRP expression not only potentially occupied the hollow capsids but also increased the concentration of available RdRP during viral assembly. Supplemented RdRPs were encoded by an arabinose-induced pBAD18-KmR-His-RdRP plasmid or by an IPTG-induced pET28b-His-RdRP. The supplementary RdRP was co-expressed with both viral RNA segments. In the absence of arabinose, AraC protein binds around the promoter (araBAD or P_{BAD}) and represses the downstream protein expression[129]. With the presence of L-arabinose, AraC proteins dimerize with the bound arabinose molecules and unveil the araBAD promoter, allowing RNA polymerase to bind the araBAD promoter and initiate the transcription of the downstream sequences for protein expression[129]. Lacking engineered arabinose transport systems (*araE* and *araFGH*), *E. coli* Rosetta 2 (DE3) cells were unable to accumulate arabinose, causing a leaking expression of the encoded RdRP from pBAD18-KmR-His-RdRP. By contrast, RdRP expression from pET28b-His-RdRP was fully active after IPTG induction, with T7 RNA polymerase expression activated under an IPTG-inducible lacUV5 promoter (i.e., λDE3 lysogen).

As expected, with the loading control GAPDH, the mild supplementary expression of RdRP from the pBAD18 vector did not significantly reinforce the overall amount of RdRP in the induced whole lysate (Fig. 6.1.1*A*). Correspondingly, the RdRP presence in the purified VLPs remained deficient and undetected (Fig. 6.1.1*B*). The detection results indicated that supplemented RdRP from the pBAD18 vector was insufficient to enhance the RdRP incorporation. However, the supplementary RdRP expression from pET28b-His-RdRP boosted RdRP expression to a similar level of codon-optimized His-RdRP expression (Fig. 6.1.1*A*), which caused the His-tagged RdRP eventually detectable in purified VLPs (Fig. 6.1.1*B*). Combined the absence of RdRP in VLPs from the co-expression of two viral RNA segments, as well as in VLPs from previous CP and RdRP co-expression[83], the RdRP presence in VLPs from RdRP supplemented co-expression of hPBV1 and hPBV2 pointed out that the presence of viral RNA and the sufficient RdRP available during assembly together determined the incorporation of RdRP in recombinant VLPs. The identity of packaged RdRP was confirmed by protein identity MS sequencing (UTHealth Proteomics Service Center). The sequencing results confirmed the existence of His-RdRP in the purified VLP sample from CsCl density gradient ultracentrifugation, whereas ribosomal proteins were still major contaminants (Table 6.1). The presence of His-RdRP in purified recombinant VLPs claimed that hPBV RdRP was encapsulated inside the virus particles.
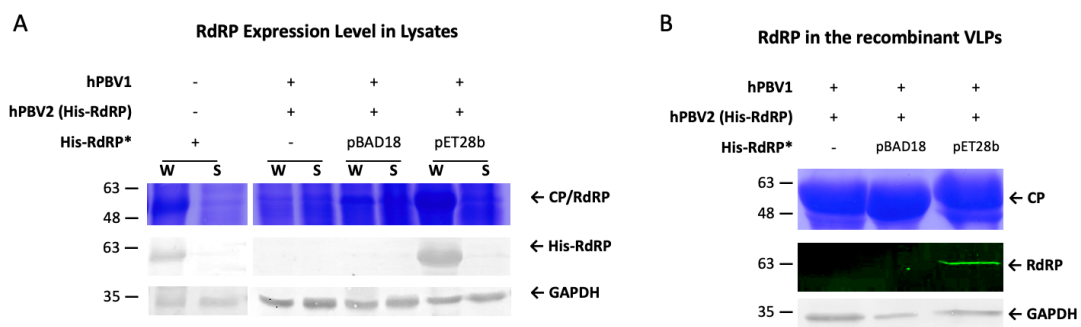
Figure 6.1.1: RdRP expression and detection in the co-expression with supplementary RdRP from different plasmids. (A) RdRP expression level with and without supplementary RdRP expression. His-tagged RdRP expressed from pET28b-His-RdRP served as the positive control. (B) His-tagged RdRP detection in the immunoprecipitated recombinant VLPs purified from ultracentrifugation. GAPDH served as the loading control. pBAD18: supplemented with pBAD18-KmR-His-RdRP expression; pET28b: supplemented with pET28b-His-RdRP expression.

Accession	Description	Score	Coverage	# Unique Peptides	# PSMs
ORF3	Capsid Protein (hPBV)	1942.84	31.16	20	975
B7UPD9	50S ribosomal protein L1 OS=Escherichia coli O127:H6 (strain E2348/69 / EPEC) OX=574521 GN=rplA PE=3 SV=1 - [RL1_ECO27]	1303.27	72.65	26	510
P0A9P0	Dihydrolipoyl dehydrogenase OS=Escherichia coli (strain K12) OX=83333 GN=lpdA PE=1 SV=2 - [DLDH_ECOLI]	517.52	50.00	30	202
A7ZIJ4	Trigger factor OS=Escherichia coli O139:H28 (strain E24377A / ETEC) OX=331111 GN=tig PE=3 SV=1 - [TIG_ECO24]	401.70	56.94	43	178
P60422	50S ribosomal protein L2 OS=Escherichia coli (strain K12) OX=83333 GN=rplB PE=1 SV=2 - [RL2_ECOLI]	383.49	67.40	24	209
B7MCS3	50S ribosomal protein L5 OS=Escherichia coli O45:K1 (strain S88 / ExPEC) OX=585035 GN=rplE PE=3 SV=1 - [RL5_ECO45]	352.25	73.18	20	184
Q1RG21	30S ribosomal protein S2 OS=Escherichia coli (strain UTI89 / UPEC) OX=364106 GN=rpsB PE=3 SV=2 - [RS2_ECOUT]	344.94	58.92	19	157
P63285	Chaperone protein ClpB OS=Escherichia coli O157:H7 OX=83334 GN=clpB PE=3 SV=1 - [CLPB_ECO57]	269.89	61.26	55	109
B1X6F7	50S ribosomal protein L6 OS=Escherichia coli (strain K12 / DH10B) OX=316385 GN=rplF PE=3 SV=1 - [RL6_ECODH]	239.54	61.58	13	126
ORF2	ORF2 (hPBV)	220.09	66.96	30	110
His-RdRP	His-RdRP (hPBV)	213.67	56.17	31	96

Table 6.1: Protein identity MS sequencing results for RdRP-overexpressed recombinant VLPs. VLP sample was purified from the co-expression of pETDuet-ΔRBS-hPBV and pET28b-His-RdRP. Score, coverage, # unique peptides, and # PSMs are the same definitions in Table 3.3.
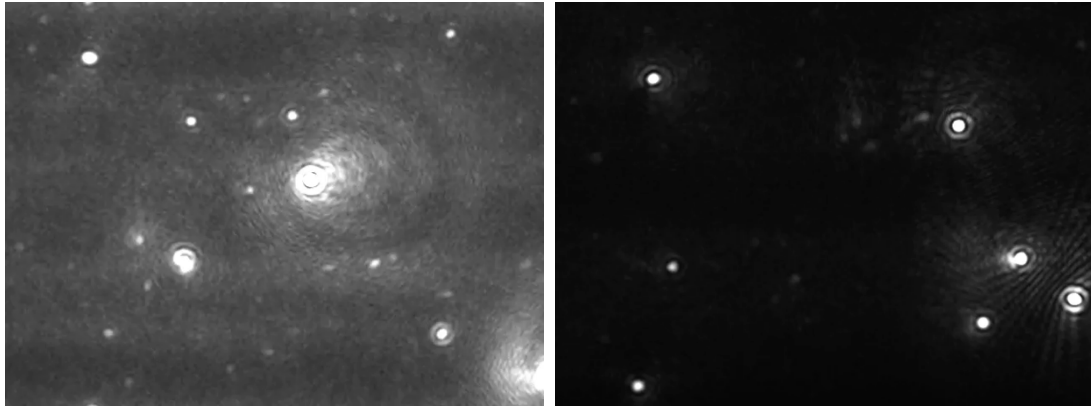
Movie 6.1: Video records of recombinant VLPs with GFP-tagged RdRP under Nanosight NS300. The left panel shows VLPs containing N-terminal GFP-tagged RdRP. The right panel displays VLPs containing C-terminal GFP-tagged RdRP with lower brightness from the source. VLP samples were in PBS buffer.

As an alternative approach to overcome the detection limitation and characterize the RdRP incorporation from hPBV1 and hPBV2 co-expression, the RdRP protein encoded by hPBV2 was fused to an N- or C-terminal GFP tag, assigning a green-fluorescent biomarker to track RdRP in the recombinant VLPs. As a high-throughput screening, Nanosight NS300 recorded and tracked the movements of recombinant VLPs at a single particle level (Movie 6.1). Under the light source of Blue Laser 488 nm, particles with fluorescent signals were observed in purified VLPs samples containing

GFP-RdRP or RdRP-GFP. Nanosight measured each particle's size by tracking its Brownian motion in the liquid suspension and calculated it by the Stokes-Einstein equation for its hydrodynamic diameter. The size distribution of recombinant hPBV particle diameter was determined as a minimum of 38 nm and an estimated mean of 174.7 nm (Fig. 6.1.2*A*). Given that the diameter of hPBV VLP was reported to be around 33-37 nm, the particle size distribution indicated potential in VLP aggregation of recombinant hPBV in PBS buffer[1]. VLP aggregation was a common issue for purified VLPs, mainly associated with buffer conditions such as pH, ion strength, and temperatures[130]. The phenomenon of particle aggregation in recombinant hPBV VLPs was also supported by the fluorescent readings measured at the single particle level (Fig. 6.1.2*B*). Larger particles (over 200 nm) were frequently measured with higher fluorescent intensities, demonstrating VLP aggregation of multiple recombinant hPBVs. In conclusion, observations of virus particles with green fluorescence under Nanosight approved the presence of RdRP in assembled recombinant VLPs. Though fluorescence signals were measured at a single particle level, the percentage of recombinant particles containing GFP-tagged RdRP was hard to quantify due to the aggregation of recombinant VLP.

To quantify the percentage of RdRP-containing recombinant VLPs, I attempted to measure the relative amount of GFP-tagged RdRP in purified VLP samples by thermal
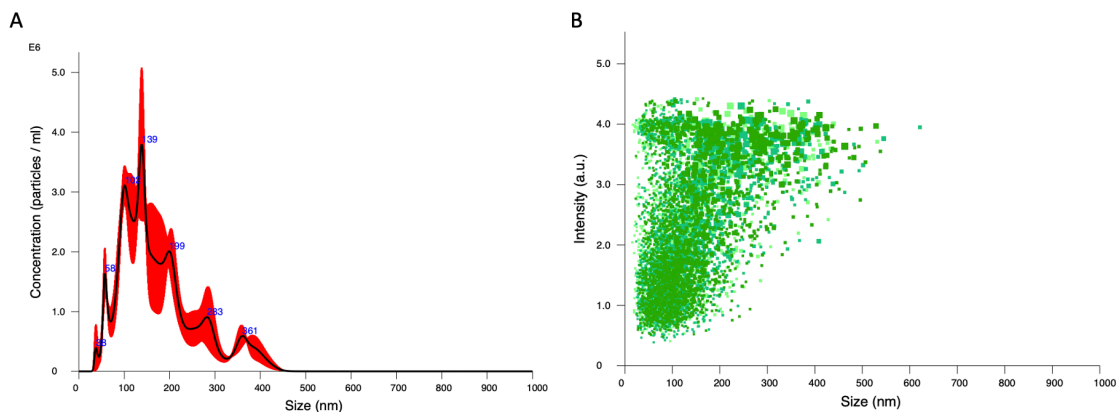
Figure 6.1.2: Size distribution and fluorescence readings of recombinant hPBV containing GFP-tagged RdRP. (A) Size distribution summaries measured and generated by Nanosight. The y-axis represents the particle concentration in 10^6 particles per ml, labeled with E6 on the top left. The x-axis represents the size of each measured particle. Standard deviations are marked as the red shade. (B) fluorescence reading of each particle. Each green dot represents the reading of each virus particle.

shift assays of GFP protein denaturation. The total protein amounts were measured by Bradford assays, and serial dilutions of VLP samples were loaded on a 96-well plate in triplicates. With increasing temperature, GFP protein unfolded and denatured, resulting in decreased fluorescence readings. The relative abundance of GFP proteins in the mixed protein sample was correlated to the relative abundance of RdRPs in assembled VLPs, where all packaged RdRP was labeled with GFP from the GFP-tagged RdRP expressed by engineered hPBV2. As a positive control, purified GFP proteins exhibited a decreasing curve with increasing temperatures, while the absolute changes of fluorescence signals were proportionally reduced in diluted samples (Fig. 6.1.3*A*). As an example of GFP-tagged protein measurement, purified VLPs from codon-optimized GFP-CP expression were measured by fluorescence signal changes with temperatures. From the test results of GFP-CP, a similar decreasing curve as purified GFP protein was observed, and the absolute changes of fluorescence signals were proportionally reduced around 1000 RFU in average from 1× to 2×, and 2× to 4× diluted samples (Fig. 6.1.3*B*). The preliminary results of GFP-CP validated the idea of measuring the relative
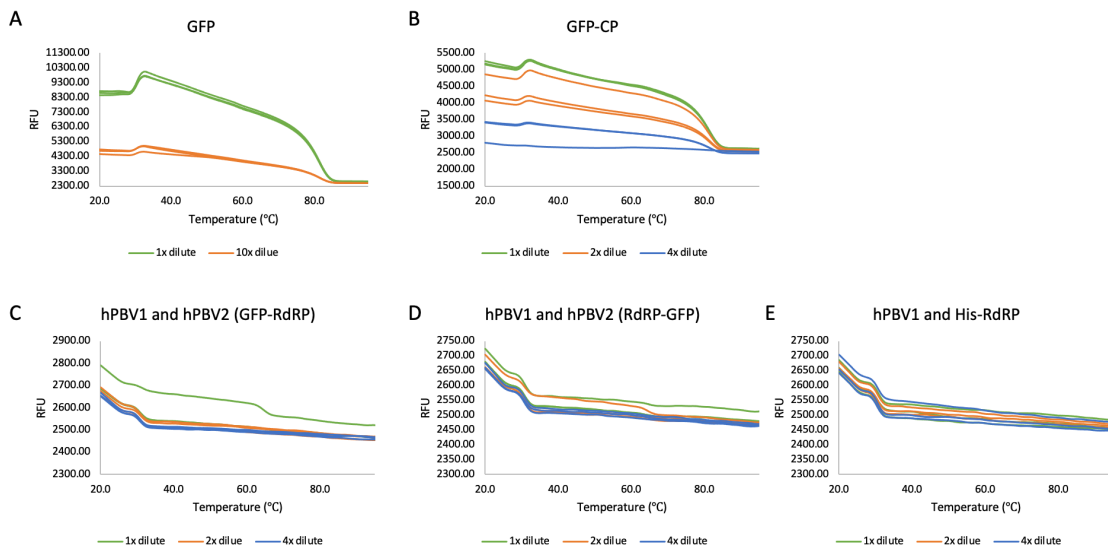
Figure 6.1.3: Thermal shift assays for relative abundance measurement of GFP and GFP-fused proteins in a mixed protein sample. (A) Free GFP protein as the positive control. (B) Purified VLPs from codon-optimized GFP-CP as an example of measuring GFP-tagged proteins in VLPs. (C) purified VLPs from the co-expression of pET28-ΔRBS-hPBV1 and pEXT20-T7-hPBV2.3 co-expression. (D) Purified VLPs from the co-expression of pET28-ΔRBS-hPBV1 and pEXT20-T7-hPBV2.4. (E) purified VLPs from pET28-ΔRBS-hPBV1 and pET28b-His-RdRP co-expression served as the negative control.

abundance of a GFP-tagged protein in a mixed protein sample by thermal shift assays. Therefore, purified VLPs containing N-terminal or C-terminal tagged RdRPs were concentrated to around 1 mg/ml for series dilutions. Negative controls were measured with ~ 1 mg/ml purified VLPs from hPBV1 and His-RdRP co-expression in the absence of GFP or GFP-tagged protein. Unfortunately, purified VLPs with GFP-tagged RdRP behaved similarly to the purified VLPs without GFP (Fig. 6.1.3*C-E*). The curves of measured fluorescence decreasing with increasing temperatures in VLPs with GFP-tagged RdRP resemble those in negative controls, which differed from those in GFP and GFP-CP measurements. Moreover, the absolute fluorescence changes for GFP-tagged RdRP were not proportional to the dilution factor. Comparing the thermal shift assay results of purified VLPs containing GFP-tagged RdRP with positive and negative controls, it concluded that the amount of incorporated GFP-tagged RdRP in the purified VLPs was below the detection limit of thermal shift assays. Thus, the packaged RdRP in purified VLPs was unable to be quantified.

6.2. The incorporation mechanism of RdRP

In previous studies, elution from Ni-NTA purification for unpackaged RdRP was examined by anti-PBV CP Western bots to characterize CP-RdRP interactions, while the purified VLPs from ultracentrifugation were detected for His-RdRP by anti-His Western blots[83]. Previous studies observed no direct interaction between the RdRP and CP was observed[83], and thus proposed that the RdRP was packaged into the viral capsids through its interactions with viral RNA segments, raising many further detailed questions about RdRP orientation, organization, and activity regulation inside hPBV particles.

In the previous section, the presence of RdRP in recombinant hPBV VLPs was illustrated by Western blot and MS sequencing from the RdRP-overexpression coupled co-expression of viral RNA segments, as well as visualization of recombinant hPBVs containing GFP-tagged RdRPs under Nanosight. Therefore, interactions between CP and RdRP were first characterized similarly to previous studies from the co-expression of viral RNA segments and supplementary RdRP. After sonication lysis and lysate clarification, the unpackaged RdRP were purified by Ni-NTA column, where some CP

proteins were co-eluted with His-RdRP (Fig. 6.2.1*A*). A cleaved CP was mainly recognized in the E1 fraction, which was later found to be caused by a contaminated protease in RNase added to the lysis buffer. The cleaved CP was N-terminal sequenced, and the cleavage was confirmed between the N44 and D45 residues. This suggested that the CP cleavage could be introduced by a protease that enzymatically cleaved between the asparagine and aspartic acid residues. The cleaved CP was no longer observed after replacing the addictive RNase with a protease-free RNase. The co-eluted CP could be the residual CP due to insufficient wash or some interaction intermediates of RdRP and CP. To further investigate the explanation of the observed co-elution of CP and RdRP, the Ni-NTA elution was loaded to CsCl density gradient ultracentrifugation. Two density lighter protein bands were observed above the VLP factions, and white protein aggregates or precipitants were obviously noticed in the middle fraction (Fig. 6.2.1*B*). To investigate the presence of RdRP-CP interaction intermediate from the protein components in each fraction, all fractions from ultracentrifugation were collected and analyzed on SDS-PAGE. VLPs were found in the lower VLP fraction with high purity, while the upper and middle fractions share a similar protein profile (Fig. 6.2.1*B*). To identify the location and possible co-localization of RdRP and CP, RdRP was detected by anti-His Western blot
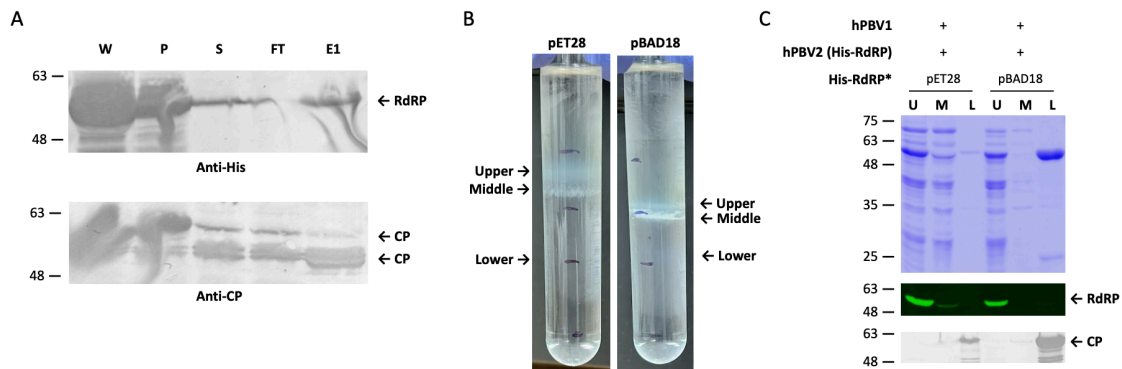
Figure 6.2.1: RdRP-CP interaction intermediates pulled down by unpackaged His-RdRP from the Ni-NTA column. (A) Ni-NTA pull-down for RdRP and CP interaction intermediates from the co-expression of pETDuet-ΔRBS-hPBV and pET28b-His-RdRP (B) CsCl density gradient ultracentrifugation results for E1 fractions from Ni-NTA column, which included possible RdRP and CP interaction intermediates. pET28: co-expression of pETDuet-ΔRBS-hPBV and pET28b-His-RdRP; pBAD18: co-expression of pETDuet-ΔRBS-hPBV and pBAD18-KmR-His-RdRP. (C) SDS-PAGE and Western blots showed that CP and RdRP were in different fractions. SDS-PAGE was displayed in the upper panel, anti-His Western blot in the middle, and anti-PBV CP Western blot at the bottom. pET28: pET28b-His-RdRP;

and CP by anti-PBV CP Western blot. The results showed that CP was only identified in lower fractions as pure VLPs, while RdRP remained in the upper fraction, indicating that RdRP from Ni-NTA elution was unpackaged. In conclusion, the co-eluted CP from the Ni-NTA column was possibly residual CP from insufficient wash before eluting.

With the hypothesis of RNA-mediated RdRP-CP interaction from the previous study and the incorporation of ORF2 found in this work, RdRP-CP interactions could only be characterized as protein-protein interactions, RNA-mediated interactions, ORF2-mediated interactions, or interactions mediated by both ORF2 and RNA. Thus, I developed an approach to immunoprecipitate the RdRP-CP interaction intermediates by anti-PBV CP polysera from different expression combinations among RdRP, CP, ORF2, and viral RNA. In immunoprecipitation, antibodies (i.e., anti-PBV CP polysera), which were anchored by the conjugated protein A/G on the magnetic beads, recognized and bound to target antigens (i.e., CPs). Immunoprecipitants (IP) were analyzed by SDS-PAGE for quality controls of antibody specificity and then by Western blots for target
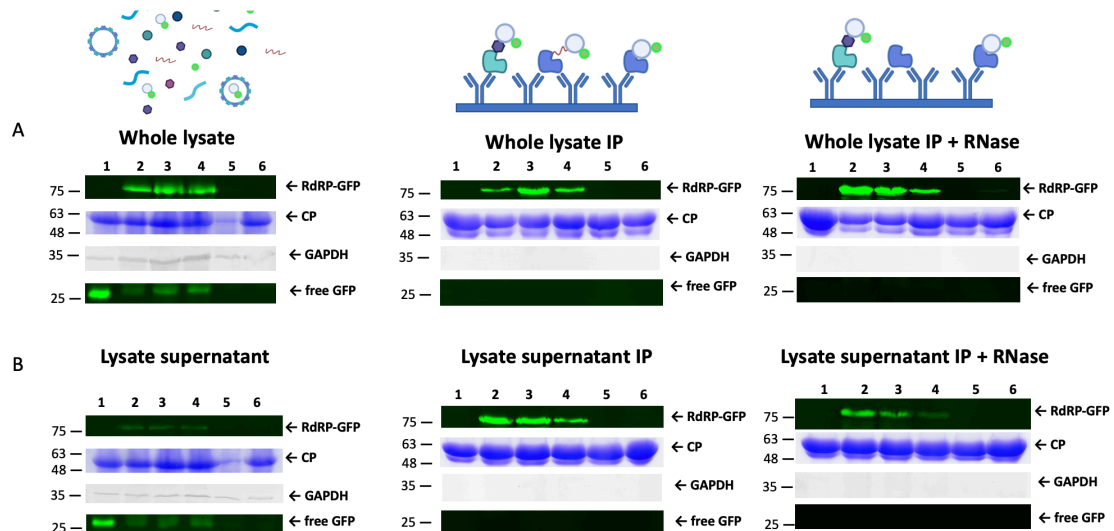
Figure 6.2.2: GFP-tagged RdRP and CP co-immunoprecipitated by the anti-PBV CP polysera. (A) Immunoprecipitation assays for whole lysate crude samples (B) Immunoprecipitation assays for the soluble lysate supernatant samples. 1: co-expression of pET28b-CP and pET19b-GFP; 2: co-expression of pET28b-CP and pET19b-RdRP-GFP; 3: co-expression of pET28-ΔRBS-hPBV1 and pET19b-RdRP-GFP; 4: co-expression of pET28-ΔRBS-hPBV1.5 and pET19b-RdRP-GFP; 5: co-expression of pET28b-CP and pEXT20-T7-hPBV2.3; 6: co-expression of pET28-ΔRBS-hPBV1 and pEXT20-T7-hPBV2.3.

protein detections. RdRP was fused with a GFP tag at C-terminus (RdRP-GFP) for detection purposes. RdRP-GFP was expressed either from engineered hPBV2 or from codon-optimized protein expression. The input samples, i.e., lysates or lysate supernatants from different expression combinations, were analyzed by SDS-PAGE and Western blots for CP and RdRP-GFP expression levels, with the detection of GAPDH as loading controls. After immunoprecipitation assays, the detections of GAPDH and free GFP in IPs presented non-specific interactions with CP from host proteins or heterologous protein expression, presumably due to random packaging.

As a negative control, immunoprecipitants from co-expression of codon-optimized CP and GFP showed that GFP alone did not interact with CP, eliminating the possibility that RdRP-GFP was immunoprecipitated by GFP-CP interactions (Lane 1 in Fig. 6.2.2). From the input lysate samples for all expression combinations analyzed in this work, RdRP-GFP protein expression was only detectable in the lysates expressed from codon-optimized RdRP-GFP, consistent with the detections of free GFP from the same sample (Fig. 6.2.2*A*). The failure of detecting RdRP-GFP from engineered hPBV2 expression was consistent and anticipated with previous co-expression results from hPBV2 and hPBV1 (Fig. 3.5.1). Due to the low expression level of RdRP-GFP from the engineered hPBV2, RdRP-GFP was not detected in the following immunoprecipitation assays (Lane 5 and 6 in Fig. 6.2.2). It explained why RdRP was missing in recombinant hPBV VLPs from the co-expression of viral RNA segments, i.e., the deficient RdRP expression from hPBV2. With a detectable RdRP-GFP expression level from codon-optimized protein expression, immunoprecipitation assays were carried out with lysates from expression combinations of RdRP-GFP with codon-optimized CP, viral RNA hPBV1, or engineered hPBV1 with ORF2 deletion (Lane 2-4 in Fig. 6.2.2). RdRP-GFP was detected in the IPs from the co-expression lysate of codon-optimized CP and RdRP-GFP, indicating direct protein-protein interactions between CP and RdRP. The conclusion was confirmed by RNase treatments. RNase digested and removed RNA molecules in IP samples, while RdRP-GFP was still found in RNase-treated IP samples. The consistent results between IPs and RNase-treated IPs eliminated the possibility of RNA-mediated RdRP-CP interactions. To investigate the possibility of ORF2-mediated RdRP-CP interactions, IP results were compared between the co-expression of RdRP-

GFP and hPBV1 and the co-expression of RdRP-GFP and engineered hPBV1 with ORF2 deletion. The engineered hPBV1 abandoned ORF2 expression by removal of sequences encoding 1-107 residues of ORF2, including the start codon for translation initiation, keeping the intact inherent RBS for CP translation located in the ORF2 C-terminus coding region (119-224 residues). With the presence or absence of ORF2, RdRP-GFP was co-immunoprecipitated with CP in IPs and RNase-treated IPs (Lane 3 and 4 in Fig. 6.2.2), indicating RdRP-CP interactions were not ORF2-mediated. All IP results from lysate supernatant samples were consistent with those from lysates. Notably, GAPDH or free GFP was not detected in all IP samples, indicating specific interactions between CP and RdRP.

To further support the conclusion that RdRP directly interacted with CP, similar immunoprecipitation assays were performed with His- RdRP. IP samples from the lysate of codon-optimized His-RdRP expression served as the negative control. In the negative control, without CP expression, no His-RdRP was non-specifically recognized by anti-PBV CP polysera (Lane 1 in Fig. 6.2.3). Consistent with previous results for RdRP expression assays, His-RdRP expressed from hPBV2 was not detectable and thus not detected in the following IPs (Lane 3 in Fig. 6.2.3). His-RdRP was detected in IPs from the co-expression of codon-optimized His-RdRP and CP, with and without RNase treatment (Lane 2 in Fig. 6.2.3). His-RdRP was also co-immunoprecipitated with CP from the lysate of RdRP supplemented viral RNA co-expression in both presence and absence of RNase (Lane 4 in Fig. 6.2.3). The consistent results between IPs with His-RdRP and IPs with RdRP-GFP firmly concluded that RdRP directly interacted CP.

RdRP-CP interactions were observed in immunoprecipitation assays from different expression combinations. However, previous results showed that CP was not
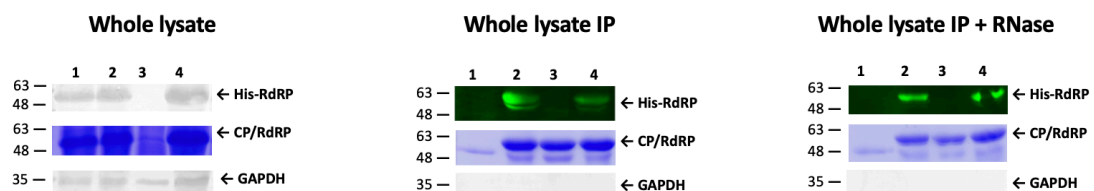
Figure 6.2.3: His-RdRP and CP co-immunoprecipitation by the anti-PBV CP polysera. 1: expression of pET28b-His-RdRP as the negative control; 2: co-expression of pET19b-CP and pET28b-His-RdRP; 3: expression of pETDuet- ΔRBS-hPBV.1; 4: co-expression of pETDuet-ΔRBS-hPBV and pET28b-His-RdRP.

pulled down by unpackaged RdRP, and RdRP was not found in VLPs assembled by CP from RdRP and CP co-expression[83]. The main difference between immunoprecipitation and Ni-NTA pull-down assays was: Ni-NTA resins anchored His-RdRP by the affinity tag on His-RdRP, while magnetic beads in immunoprecipitation anchored CP by anti-PBV CP polysera and the conjugated protein A/G on the beads. The Ni-NTA resins anchored His-RdRP with rigidity and thus exposed limited orientations to interact with CP. By contrast, CP anchored by magnetic beads through protein A/G and antibodies allowed more flexibility for different interaction orientations. Another possible explanation was that the N-terminus of RdRP was responsible for RdRP-CP interactions, which were hindered by Ni-NTA resins in the Ni-NTA pull-down assay.

6.3. Summary

The RdRP of hPBV was found in capsids, and specific protein-protein interactions were observed between CP and RdRP. The incorporation of RdRP in recombinant hPBV VLPs determines by both the accessible RdRP concentration and the presence of viral RNA segments during viral assembly. Less relative abundance of RdRP and CP was detected in the purified recombinant VLPs than in immunoprecipitation assays, implying that not all RdRP-CP interaction intermediates were assembled into recombinant VLPs. Given that the presence of viral RNA was not required for RdRP-CP interactions but required for RdRP-incorporated recombinant VLPs, I proposed that the proper interactions between viral RNA and RdRP served as a quality control step to package RdRP in the assembled recombinant VLPs. It explained that increasing the available RdRP concentration during viral assembly by supplementary RdRP expression pushed the reaction kinetics of RNA-RdRP interaction towards the yield of RdRP-incorporated recombinant VLPs.

Further studies on RdRP-CP interaction could be revealed by eluting the RdRP-CP complex from immunoprecipitation assays and analyzing their structures under cryo-EM. The proposed quality control step in viral assembly, i.e., RNA-RdRP interactions, could be unveiled by the structure comparison of RdRP-CP complexes in the presence and absence of viral RNA.

<div align="right">

Chapter 7

</div>

The potential natural host of hPBV

The actual host of PBV remained controversial. The theoretical host has been proposed to be eukaryotic cells, bacteria, or mitochondria. Eukaryotic host theory was supported by the effective interspecies transmission of PBV and the observation that no specific pathogenetic bacteria was shared among PBV-positive samples[57,63–68]. The mitochondrial host theory was proposed based on the alternative codon usage of invertebrate mitochondria observed in some PBV-like sequences[74–79]. The bacteriophage nature explained the detection of PBV in a wide range of animal species and environmental samples and the detection of multiple PBV strains in individuals[48,60,71,73]. Given that no shared bacteria was identified as the co-infection host in PBV-positive samples, PBV could potentially infect probiotics, which eventually damaged the immunity system due to imbalanced microbiota[68,131]. Microbiota is in alliance with the innate immune system, while autoimmune and inflammatory disorders were observed more frequently with affected microbiota[132]. It also explained that, in clinics, PBV prevalence was proposed as a biomarker of immunosuppression[62]. In this work, the production of recombinant hPBV VLPs from viral RNA expression in *E. coli* Rosetta 2 (DE3) cells supported the prokaryotic host theory. Thus, the expression host, *E. coli*, was examined as a potential host of PBV.

7.1. *E. coli* was not the natural host of PBV

     E. coli served as the expression host in the recombinant hPBV expression and thus was first analyzed whether it was a potential host of PBV. Viral infections require specific interactions between the viral capsids and host surface proteins[133,134]. To unveil whether *E. coli* was a potential host of PBV, the lysate of *E. coli* cells was immunoprecipitated by antibody-bound recombinant hPBV VLPs and immunoprecipitants (IP) were profiled by SDS-PAGE. As a negative control, *E. coli* lysates were immunoprecipitated against anti-PBV CP polysera alone to validate the results by antibody specificity. As the negative control results showed, besides the protein band assigned to the antibody heavy chain at around 50 kDa, no other protein band on SDS-PAGE was observed as a non-specific target of anti-PBV CP polysera (Fig. 7.1). Since recombinant PBVs randomly packaged host proteins, mainly ribosomal proteins, the protein profile of immunoprecipitants from purified recombinant VLPs served as test
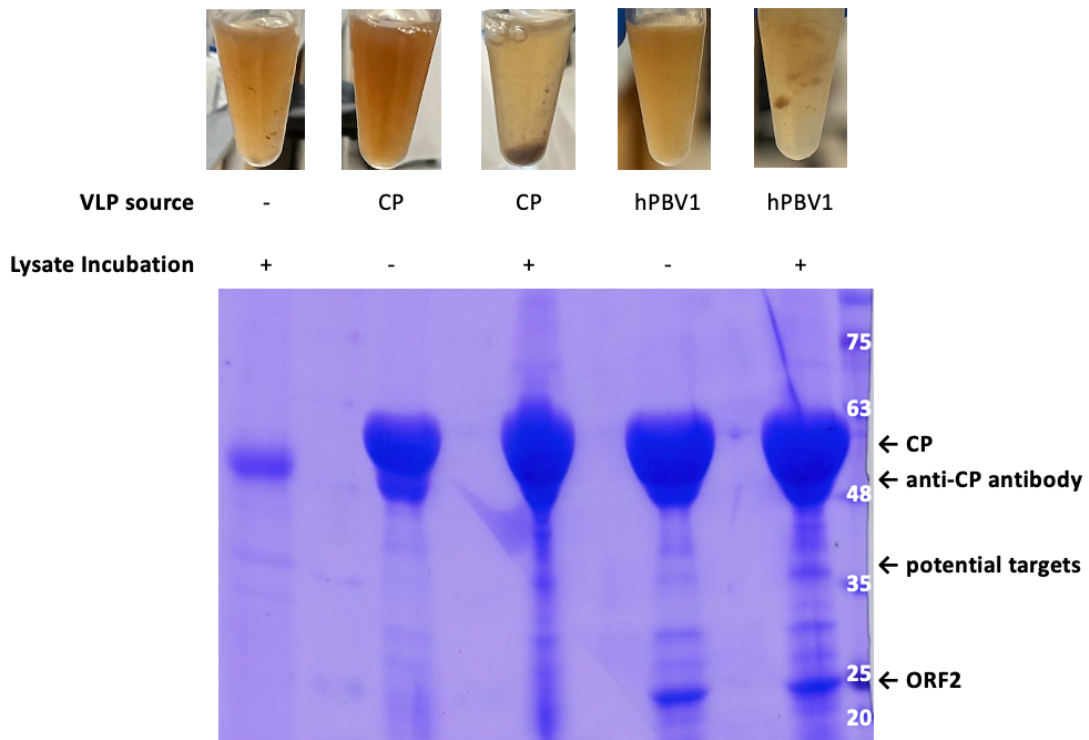
Figure 7.1: Co-immunoprecipitated *E. coli* host proteins with recombinant hPBV VLPs from *E. coli* lysate. The upper panel showed the magnetic beads after immunoprecipitation incubation, where aggregations were observed in some of the tubes. The lower panel provided the SDS-PAGE profile for each corresponding immunoprecipitated sample. The amount of immunoprecipitated CP served as the loading control. CP: purified recombinant VLPs from codon-optimized CP expression; hPBV1: purified recombinant VLPs from hPBV1 expression.

background (Fig. 7.1). To eliminate variation from co-expression of hPBV1 and hPBV2, which produced heterogeneous recombinant VLPs with and without RdRP incorporation, recombinant VLPs used in host protein pull-down assays were yielded from codon-optimized CP or hPBV1 expression. The immunoprecipitated recombinant hPBV VLPs were then incubated with *E. coli* lysates to pull down the host proteins that specifically interact with the hPBV capsids. In the protein profiles of host proteins pulled down by hPBV VLPs, a relatively increased amount of proteins at around 37 kDa was observed on SDS-PAGE, which could be the potential host protein targeted by hPBV (Figure 7.1). An interesting observation along the immunoprecipitation assays was that aggregation of magnetic beards was usually observed after incubating the immunoprecipitated recombinant hPBV with *E. coli* lysate, possibly due to hPBV-host protein interactions.

To unveil the identities of *E. coli* host proteins interacting with hPBV, co-immunoprecipitated host proteins were analyzed by mass spectrometry (MS) sequencing (UTHealth Proteomics Service Center). Similar to protein profile comparisons by SDS-PAGE, immunoprecipitated *E. coli* lysates against anti-PBV CP polysera alone served as a negative control; immunoprecipitated recombinant VLPs from hPBV1 expression as background; and host proteins pulled down by immunoprecipitated VLPs as a test sample. MS sequencing was performed for the negative control sample, the background sample, and the test sample. In sequencing analyses, comparisons among the three samples were only performed for identified proteins with a sequence coverage percentage higher than 20% or the identified sequence number (#PSM) higher than 10. The set threshold eliminated most identified host proteins with non-specific interactions with recombinant VLPs.

E. coli host proteins only identified in the test sample but absent in the negative control or background were summarized in Table 7.1.1. Interestingly, an *E. coli* polymerase was pulled down, and correspondingly a ribosomal protein L17. They were possibly pulled down by the RNA molecules that interacted with hPBV capsids. Among other identified host proteins, 5-methyltetrahydropteroyltriglutamate-homocysteine methyltransferase was a methyltransferase in methionine formation; 2,3,4,5-tetrahydropyridine-2,6-dicarboxylate N-succinyltransferase was an acyltransferase in lysine biosynthesis, biosynthetic arginine decarboxylase catalyzed biosynthesis from

arginine to agmatine; phosphoglycerate kinase involved in glycolysis and located in the cytoplasm; and chaperone protein DnaJ protect proteins from irreversible aggregation. The molecular weights of phosphoglycerate kinase and chaperone protein DnaJ were close to the proposed target host protein at around 37 kDa in the protein profile analysis on SDS-PAGE. It indicated that the ~ 37 kDa co-immunoprecipitated host protein was associated with protein aggregation caused by cell lysis or recombinant hPBVs.

E. coli host protein with increased relative abundance in the test sample compared with the negative control or background were summarized in Table 7.1.2. Among the identified potential targets, 3-oxoacyl-[acyl-carrier-protein] synthase 2 involved in type II fatty acid elongation cycle found in the cytoplasm; cysteine desulfurase lscS delivered suffer to Fe-S cluster synthesis; L-threonine dehydratase biosynthetic IlvA catalyzed dehydration of threonine into alpha-ketobutarate and ammonia; chloramphenicol acetyltransferase was the effector of chloramphenicol resistance, which came from the pLysS plasmid to express eukaryotic tRNAs in Rosetta cells; D-tagatose-1,6-bisphosphate aldolase subunit GatZ involved in the catabolism of galactitol. All of them were involved in cellular biosynthesis found in the cytosol.

In conclusion, none of the identified potential host targets from MS sequencing analyses was a membrane-associated protein on the E. coli surface. The host protein candidate at around 37 kDa in the protein profile analysis on SDS-PAGE was likely to be a chaperon protein due to the protein aggregation from lysis or immunoprecipitation assays. Therefore, E. coli was not a potential host of PBV, lacking a host protein target on the E. coli surface to anchor PBV capsids. This work provided an approach for future studies to unveil the natural host of PBV, where the E. coli lysate used in immunoprecipitation assays could be replaced by cell lysate of other bacteria species or a mixed biological sample like feces.

Protein Identity MS Sequencing Comparison

Accession	Description	M.W. (kDa)	VLP + Lysate IP			Anti-PBV CP IP control			VLP IP control		
			Score	Coverage	# PSMs	Score	Coverage	# PSMs	Score	Coverage	# PSMs
C5A0S7	DNA-directed RNA polymerase subunit beta	150.5	49.84	23.85	32						
Q1R638	50S ribosomal protein L17	14.4	25.02	52.76	17						
P25665	5-methyltetrahydropteroyltriglutamate-homocysteine methyltransferase	84.6	20.76	10.09	8						
B1LGW7	2,3,4,5-tetrahydropyridine-2,6-dicarboxylate N-succinyltransferase	29.8	20.63	33.94	10						
B1LDE8	Biosynthetic arginine decarboxylase	71.1	18.99	17.09	10						
A8A466	Phosphoglycerate kinase	41.1	17.54	26.10	10						
B5YYA8	Chaperone protein DnaJ	41.0	15.24	21.01	10						

Table 7.1.1: *E.coli* **host proteins pulled down by immunoprecipitated recombinant VLPs.** VLP + lysate IP: *E. coli* host proteins pulled down by immunoprecipitated recombinant hPBV VLPs; Anti-PBV CP IP control: negative control of non-specific interaction between host proteins and anti-PBV CP polysera; VLP IP control: immunoprecipitated VLPs as detection background. Identified proteins with coverage > 20% or # PSMs > 10 were analyzed. Score: Displays the protein score, which is the sum of the scores of the individual peptides. Coverage: Displays by default the percentage of the protein sequence covered by identified peptides. # PSMs: Displays the total number of identified peptide sequences (peptide spectrum matches) for the protein, including those redundantly identified.

Protein Identity MS Sequencing Comparison

Accession	Description	M.W. (kDa)	VLP + Lysate IP			Anti-PBV CP IP control			VLP IP control		
			Score	Coverage	# PSMs	Score	Coverage	# PSMs	Score	Coverage	# PSMs
P0AAI5	3-oxoacyl-[acyl-carrier-protein] synthase 2	43.0	29.12	16.95	15	14.54	15.25	7	8.10	11.38	4
B7M7N3	Cysteine desulfurase IscS	45.1	28.54	32.18	15	16.35	21.04	10			
P04968	L-threonine dehydratase biosynthetic IlvA	56.2	23.58	31.71	16	4.08	3.31	1			
P62577	Chloramphenicol acetyltransferase	25.6	15.69	16.44	12	11.23	15.53	7	0.00	15.98	4
A7ZNR4	D-tagatose-1,6-bisphosphate aldolase subunit GatZ	47.1	40.74	24.76	23	34.86	26.43	20	9.59	10.24	4

Table 7.1.2: *E. coli* **host proteins with higher abundance in co-immunoprecipitants than the negative control and the immunoprecipitated VLP background.** VLP + lysate IP: *E. coli* host proteins pulled down by immunoprecipitated recombinant hPBV VLPs. Anti-PBV CP IP control: negative control of non-specific interaction between host proteins and anti-PBV CP polysera; VLP IP control: immunoprecipitated VLPs as detection background. Score, coverage, and # PSMs are the same definitions in Table 7.1.1.

<div align="right">

Chapter 8

</div>

Major Conclusions

In this work, the expression system of recombinant hPBV was established by viral RNA expression in *E. coli* Rosetta 2 (DE3) cells—the expression of individual hPBV viral RNA segments successfully translated into all predicted ORFs. The functionality of all predicted RBSs indicated that hPBV was a prokaryotic virus infecting some bacteria species in the human microbiome. This explained the detection of PBV in feces from a wide range of animal species and multiple PBV strains detected in individual samples[48,60,73]. The bacteriophage nature also demonstrated that PBV infection played a synergetic role in many diseases and was proposed as an immunosuppression biomarker in clinics[62]. Gut microbiome imbalance (dysbiosis) could affect the immune system and cause diarrhea and gastroenteritis, inflammation, diabetes, atopy, and cancer[135,136]. In addition, the yield of recombinant hPBV from the non-natural host *E. coli* Rosetta strains by the co-expression of both viral RNA segments not only provided a model for hPBV studies but also approved the hypothesis of hPBV as a prokaryotic virus.

Quality control experiments of the recombinant hPBV expression system showed that most HDV ribozyme sequences were efficiently self-removed. The only ineffective self-cleavage activity of the HDV ribozyme sequence was observed at the end of hPBV1 expressed from pETDuet-ΔRBS-hPBV, indicating the presence of a long RNA transcript as hPBV1-HDV-T7-hPBV2. Though RdRP translation from hPBV2 was suppressed during its co-expression with hPBV1, the expressed RdRP was remarked by its activity in viral RNA synthesis and the subsequent increase in viral protein expression. Further optimization of the hPBV expression system was proposed through an optimized

ribozyme sequence at the 3'-end of viral RNA after synthesizing transcripts. Alternatively, the RdRP expression from hPBV2 could be improved by investigating the translation regulation mechanism of hPBV RNA segments by manipulating the upstream inherent RBS in viral RNA sequences. RdRP expression suppression during co-expressions could also be related to the expression host, since *E. coli* was not the actual host of PBV.

In recombinant hPBV VLPs, ORF2, RdRP, and viral RNA segments were packaged in assembled virus particles. The small 39-amino-acid ORF1 was expressed but not found in the recombinant particles. Major conclusions for the packaging of ORF2, RdRP, and viral RNA segments were as follows:

Though hPBV capsids randomly packaged RNA molecules by their positively charged N-terminus, viral RNA segments were specifically packaged into recombinant VLPs. Each viral RNA segment (hPBV1 and hPBV2) could be packaged separately and independently of each other. The selective packaging of viral RNA segments required the presence of packaging signal sequences at the 5'- and 3'-untranslated regions. However, it remained unknown if each recombinant hPBV particle encapsulated two RNA segments in an equimolar ratio or packaged one of the segments at a higher copy number. Answers to this question required analyses of the encapsulated RNAs at the resolution of a single recombinant virus particle. In addition, since each RNA segment could be packaged separately, future experiments could also investigate if any recombinant VLPs incorporated only one RNA segment.

The 224-amino-acid ORF2 was poorly folded and only soluble in the presence of CP during translation. ORF2 protein comprised a disordered N-terminal domain containing seven conserved ExxRxNxxxE motifs and a hydrophobic C-terminal domain proposed to interact with viral RNA. The disordered N-terminal domain of ORF2 interacted with CP resulting in the encapsulation of ORF2, which could be entirely disrupted by removing all repeated ExxRxNxxxE motifs in the first 102 residues. Packaged ORF2 was organized as a spherical protein layer underneath the viral capsids. The relative abundance of CP and ORF2 in the purified VLPs showed that each assembled virus particle encapsidated around 9.12 to 11.16 copies of ORF2 on average, with a maximum of around 20.64 copies. Based on the calculated copy number of

packaged ORF2, ORF2 proteins were proposed to distribute around either the twelve 5-fold vertices or the twenty 3-fold vertices. Future investigation of ORF2 and viral RNA interactions would demonstrate the predicted RNA-interacting activity of the ORF2 C-terminal domain.

RdRP interacted with CP by direct protein-protein interactions. However, the packaging of RdRP in recombinant VLPs determines by the presence of both viral RNA segments and the amount of available RdRP during assembly. The RdRP expression level from merely hPBV2 was insufficient for RdRP to be incorporated into recombinant VLPs, whereas a supplemented RdRP expression was required. Since the presence of viral RNA was not required for RdRP-CP interactions but required for RdRP-incorporated recombinant VLPs, the proper interactions between viral RNA and RdRP were proposed to be the quality control step to assemble RdRP-packaged recombinant VLPs. The RdRP-CP interactions could be structurally characterized by solving the RdRP-CP complex structure under cryo-EM.

Last but not least, the expression host *E. coli* was not the natural host of PBV. Though several candidates were identified from the pull-down assays of *E. coli* host proteins against immunoprecipitated recombinant hPBV by SDS-PAGE and MS sequencing, none of them was located on the cell surface, indicating no surface target in *E coli* was accessible for PBVs. Thus, it concluded that *E. coli* was not the actual host of PBV. The host protein pull-down analyses by recombinant hPBV provided an approach for future studies to unveil the natural host of PBV. As *E. coli* was not the actual host of PBV, this work also provided an example of expressing a novel dsRNA virus in an unnatural host system before unveiling its real host.

Chapter 9

Future Work

Many future experiments have been proposed in this work: (1) optimizing ribozyme self-cleavage activity by testing different ribozyme sequences at the 3'-end of viral RNA segments; (2) investigating translation regulation mechanism of the viral RNA by manipulating the upstream inherent RBS of each ORF; (3) characterizing the interactions between viral RNA and RdRP via internal deletions of viral RNA; (4) probing the RNA sequencing in each recombinant hPBV particle to illustrate if the presence of each RNA segments at single particle resolution; (5) characterizing ORF2 C-terminal domain for its RNA binding activity; (6) resolving the *in situ* ORF2 structures from ORF2-incorporated recombinant hPBV by cryo-EM; (7) structural studies in RdRP-CP interactions through resolving the structure of RdRP-CP complex by cryo-EM; (8) unveiling the actual host of PBV through pull-down assays of lysate from different bacteria against immunoprecipitated recombinant hPBV VLPs.

Among the proposed experiments, structural studies rely on the homogeneity of a purified protein sample. Therefore an *in vitro* assembly system and a particle sorting system was provided and presented here for future works.

9.1. CP disassembly and re-assembly for hPBV assembly *in* vitro

I attempted to establish a CP disassembly and re-assembly *in vitro* system for future protein-protein or protein-RNA interactions *in vitro* studies during viral assembly. CP of hPBV automatically self-assembled into capsids after expression. Features of CP

disassembly were analyzed by CD spectrometry. The assembled CPs were expressed from pET19b-CP and purified by CsCl density gradient ultracentrifuge. The assembled CP was analyzed by CD spectrometry with increasing temperatures (Fig. 9.1.1*A*). Based on the spectra, the melting temperature of assembled CP was estimated to be around 65 °C. After interpreting the CD spectra into secondary structure composition, increases in α-helix and turns while a complete loss in parallel β-strand was noticed around the melting temperature (Fig. 9.1.1*C*). It implied that the disassembly of CP was marked by denature of the parallel β-sheets, which were primarily found in the P (projecting) domain of CP (Fig. 1.2.3.2)[37]. The loss of parallel β-sheets in the P domain interrupted the intricate interface of the dimer, allowing the disassociation of the CP dimer. The denatured CP was characterized in PBS buffer containing 8 M urea (Fig. 9.1.1*B*). Secondary structure analysis from the CD spectra showed that the denatured CP was completely unfolded, lacking remarkable features of α-helix and β-sheet (Fig. 9.1.1*D*). Thus, it was concluded that assembled VLPs could be completely denatured by 8 M urea. However, the VLP disassembly via increasing temperatures or in 8 M urea was not reversible, as protein aggregates were observed at the melting temperature. In addition,
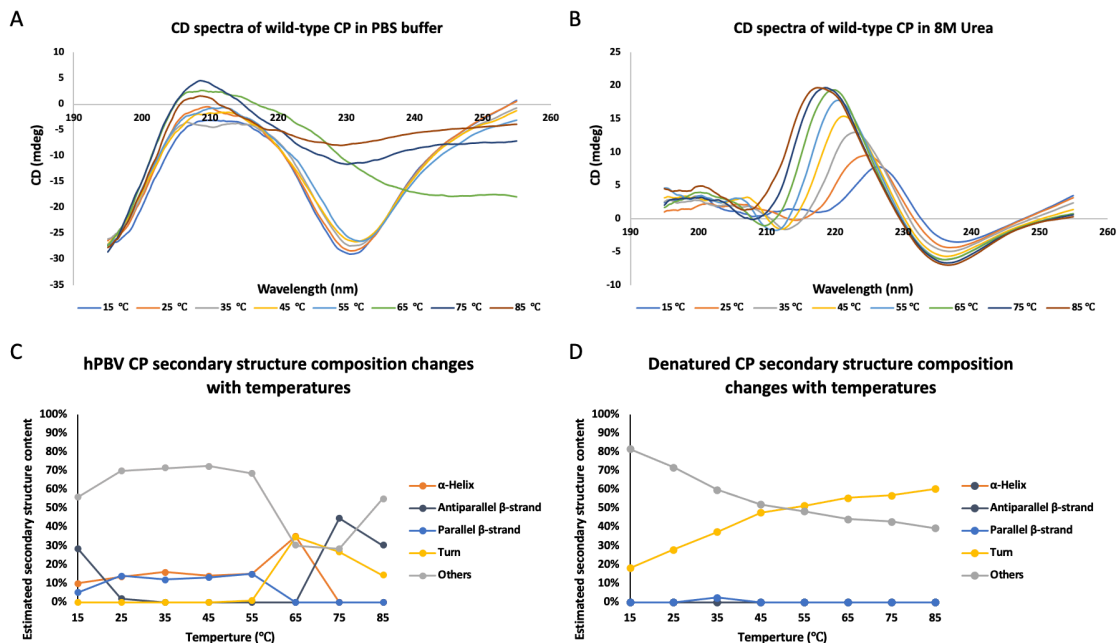

Figure 9.1.1: CD measurement of assembled and denatured CP. (A) Temperature scan of CD measurement of assembled CP in PBS buffer. (B) Temperature scan of CD measurement of denatured CP in PBS buffer with 8 M urea. (C) Estimated secondary structure composition of assembled CP from spectra in (A). (D) Estimated secondary structure composition of denatured CP from spectra in (B).

VLP assembly usually involves a domain swap in capsid proteins[137]. In PBV, the domain swap was observed at the intricate interface of hPBV CP dimer[37,38]. Based on the secondary structure composition analysis, the intricate interface in the P domain was likely disrupted at 65 °C and completely disrupted in 8 M urea, indicating an irreversible disassembly.

In previous studies, disassembly and reassembly of hPBV capsids were characterized as a reversible reaction by dialysis in mild buffers[37]. According to the literature, VLP disassembly was performed by dialysis in buffer A (50 mM Na_2CO_3 pH 10 and 1 M NaCl) at 20 °C for 12 hours and reassembly in buffer B (25 mM Tris-HCl pH 7.5, 50 mM NaCl)[37]. Repeated experiments were carried out in the same condition. However, the disassembled CP formed white aggregates after dialysis in buffer A. The solubility issue of the disassembled CP was solved by diluting the input VLP sample to around 0.5 mg/ml. Under TEM, disassembled CP was observed in either rod or diamond shape, around 10-20 nm in length, corresponding to the dimer or the tetramer tile (Fig. 10.1.2*B*). The observation of the VLP building blocks, CP dimers and tetramers, indicated a successful repeat experiment for VLP disassembly. However, the disassembled CPs were mainly precipitated after dialysis in buffer B, where less than 10% of the proteins remained soluble after reassembly. Eventually, a few reassembled VLP particles were obtained and analyzed under the TEM. As the TEM image showed, protein aggregates and reassembled VLPs were observed. Compared with purified recombinant VLPs, some of the reassembled VLPs lacked the "smooth" texture of capsid surface as poorly assembled. I hypothesized that the quality of reassembly was related to buffer conditions. To further characterize the VLP reassembly, the disassembled CPs
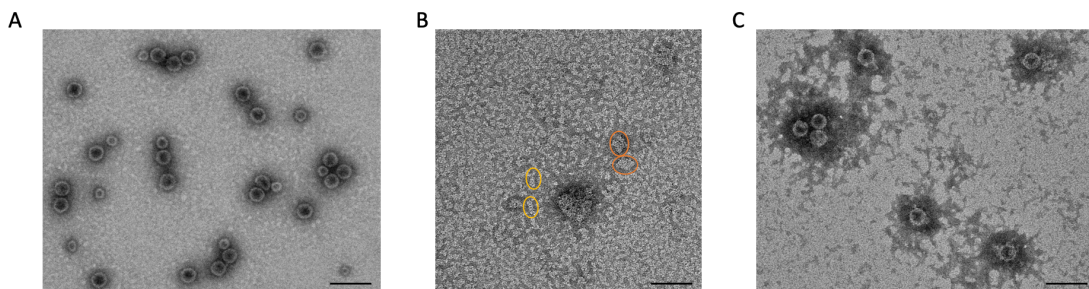

Figure 10.1.2: TEM images of disassembled and reassembled CPs. (A) Purified CP VLP under TEM. (B) Disassembled CP after dialysis in buffer A at 4 °C overnight. (C) Reassembled VLP after dialysis in PBS.

were gradually diluted by dripping 1 ml PBS buffer per day. Protein aggregates were observable after adding 4 ml PBS buffer, with pH around 9.5 and a salt concentration around 320 mM (Fig. 10.1.3). Since the buffer pH was not much dropped, the results indicated salt concentration might play a more critical role in facilitating CP reassembly. Future experiments could explore more factors in reassembly that affected VLP reassembly quality, such as temperatures, buffer compositions, and the presence of viral RNA.

PBS	Salt conc. (M)
0 ml	1.000
1 ml	0.575
2 ml	0.433
3 ml	0.363
4 ml	0.320
5 ml	0.292
6 ml	0.271
7 ml	0.256
8 ml	0.244
9 ml	0.235
10 ml	0.227
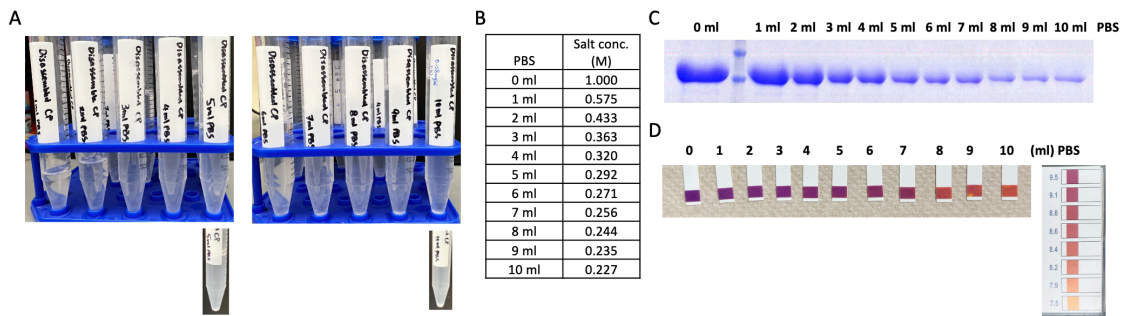
Figure 9.1.3: CP reassembly in PBS dilution. (A) Reassembly assay in PBS dilution. (B) Chart for final salt concentration after PBS dilution. (C) SDS-PAGE for final protein concentration after PBS dilution. (D) Final pH after PBS dilution.

9.2. ORF2 orientation in VLPs and application for VLP sorting

One of the main obstacles of viral vector applications was to differentiate the packaged particles from the empty capsids. In addition, sorting ORF2-incorporated VLPs could improve the sample homogeneity for cryo-EM analysis and eventually facilitate solving the *in situ* ORF2 protein structure. Since ORF2 was incorporated in the recombinant hPBVs, an attempt to sort ORF2-incorporated VLPs was performed by immunoprecipitating the recombinant VLPs against commercial anti-His antibody, which recognized the N-terminal His tag on ORF2 protein. Recombinant VLPs were purified from the co-expression of hPBV1 and His-ORF2. Surprisingly, both CP and ORF2 were found in the immunoprecipitants from purified VLPs (Fig. 9.2). Negative controls were then performed by immunoprecipitation assays with only purified VLPs or only antibodies. As expected, no CP or ORF2 was detected in the negative control of

immunoprecipitants (IPs) from the antibody-only experiment. Few CPs but no ORF2 were detected in the negative control of immunoprecipitants (IPs) from purified VLP without antibodies. The promising results indicated that sorting ORF2-incorporated VLPs was feasible. As the N-terminal domain of ORF2 interacted with CP, sorting ORF2-incorporated VLP by the N-terminal His tag on ORF2 suggested that the N-terminus of ORF2 pointed exterior, making it accessible for anti-His antibody.

Future experiments could examine the orientation of the ORF2 C-terminus with a C-terminal Strep tag on the ORF2 protein. Immunoprecipitating ORF2-packaged VLPs by anti-Strep antibody will illustrate the direction of ORF2 C-terminus, i.e., pointing interior or exterior correlated with the absence or presence of ORF2 in IPs (Fig. 10.2*A*). Furthermore, a protein of interest could be inserted or fused to ORF2. The sorting system for ORF2-incorporated recombinant VLPs would eventually sort the recombinant VLPs packaged with the protein of interest.

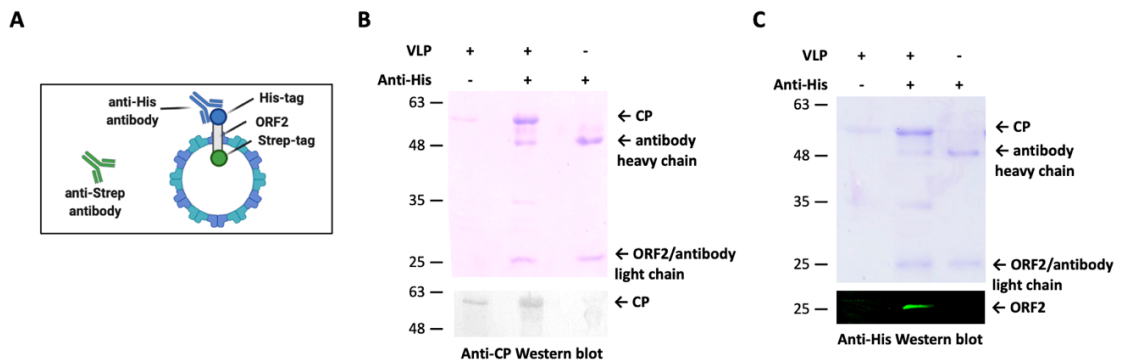**Figure 9.2: VLP sorting via the orientation of ORF2 in capsids.** (A) The scheme of VLP sorting by immunoprecipitation. ORF2 protein was fused with His-tag on the N-termini and Strep-tag on the C-termini. (B) SDS-PAGE and Western blot showed His-ORF2 incorporated VLPs were immunoprecipitated after anti-His immunoprecipitation. (D) His-tagged ORF2 was only detected in the immunoprecipitated VLPs.

References

1. Delmas, B., Attoui, H., Ghosh, S., Malik, Y.S., Mundt, E., Vakharia, V.N., and ICTV Report Consortium (2019). ICTV virus taxonomy profile: Picobirnaviridae. Journal of General Virology *100*, 133–134. 10.1099/jgv.0.001186.

2. Delmas, B., Attoui, H., Ghosh, S., Malik, Y.S., Mundt, E., Vakharia, V.N., and ICTV Report Consortium (2019). ICTV virus taxonomy profile: Birnaviridae. Journal of General Virology *100*, 5–6. 10.1099/jgv.0.001185.

3. Kotta-Loizou, I., Coutts, R.H.A., and ICTV Report Consortium (2022). ICTV Virus Taxonomy Profile: Polymycoviridae 2022: This article is part of the ICTV Profiles collection. Journal of General Virology *103*. 10.1099/jgv.0.001747.

4. Kotta-Loizou, I., Castón, J.R., Coutts, R.H.A., Hillman, B.I., Jiang, D., Kim, D.-H., Moriyama, H., Suzuki, N., and ICTV Report Consortium (2020). ICTV Virus Taxonomy Profile: Chrysoviridae. Journal of General Virology *101*, 143–144. 10.1099/jgv.0.001383.

5. Poranen, M.M., Mäntynen, S., and ICTV Report Consortium (2017). ICTV Virus Taxonomy Profile: Cystoviridae. Journal of General Virology *98*, 2423–2424. 10.1099/jgv.0.000928.

6. Sato, Y., Miyazaki, N., Kanematsu, S., Xie, J., Ghabrial, S.A., Hillman, B.I., Suzuki, N., and ICTV Report Consortium (2019). ICTV Virus Taxonomy Profile: Megabirnaviridae. Journal of General Virology *100*, 1269–1270. 10.1099/jgv.0.001297.

7. Vainio, E.J., Chiba, S., Ghabrial, S.A., Maiss, E., Roossinck, M., Sabanadzovic, S., Suzuki, N., Xie, J., Nibert, M., and ICTV Report Consortium (2018). ICTV Virus Taxonomy Profile: Partitiviridae. Journal of General Virology *99*, 17–18. 10.1099/jgv.0.000985.

8. Chiba, S., Castón, J.R., Ghabrial, S.A., Suzuki, N., and ICTV Report Consortium (2018). ICTV Virus Taxonomy Profile: Quadriviridae. Journal of General Virology *99*, 1480–1481. 10.1099/jgv.0.001152.

9. Matthijnssens, J., Attoui, H., Bányai, K., Brussaard, C.P.D., Danthi, P., del Vas, M., Dermody, T.S., Duncan, R., Fāng (方勤), Q., Johne, R., et al. (2022). ICTV Virus Taxonomy Profile: Sedoreoviridae 2022. Journal of General Virology *103*. 10.1099/jgv.0.001782.

10. Matthijnssens, J., Attoui, H., Bányai, K., Brussaard, C.P.D., Danthi, P., del Vas, M., Dermody, T.S., Duncan, R., Fāng, Q., Johne, R., et al. (2022). ICTV Virus

Taxonomy Profile: Spinareoviridae 2022. Journal of General Virology *103*. 10.1099/jgv.0.001781.

11.   Chen, Y.G., and Hur, S. (2022). Cellular origins of dsRNA, their recognition and consequences. Nat Rev Mol Cell Biol *23*, 286–301. 10.1038/s41580-021-00430-1.

12.   Ahlquist, P. (2006). Parallels among positive-strand RNA viruses, reverse-transcribing viruses and double-stranded RNA viruses. Nature Reviews Microbiology *4*, 371. 10.1038/nrmicro1389.

13.   Mertens, P. (2004). The dsRNA viruses. Virus Research *101*, 3–13. 10.1016/j.virusres.2003.12.002.

14.   Pan, M., Alvarez-Cabrera, A.L., Kang, J.S., Wang, L., Fan, C., and Zhou, Z.H. (2021). Asymmetric reconstruction of mammalian reovirus reveals interactions among RNA, transcriptional factor μ2 and capsid proteins. Nat Commun *12*, 4176. 10.1038/s41467-021-24455-4.

15.   Yu, X., Ge, P., Jiang, J., Atanasov, I., and Zhou, Z.H. (2011). Atomic Model of CPV Reveals the Mechanism Used by This Single-Shelled Virus to Economically Carry Out Functions Conserved in Multishelled Reoviruses. Structure *19*, 652–661. 10.1016/j.str.2011.03.003.

16.   Miyazaki, N., Uehara-Ichiki, T., Xing, L., Bergman, L., Higashiura, A., Nakagawa, A., Omura, T., and Cheng, R.H. (2008). Structural Evolution of *Reoviridae* Revealed by *Oryzavirus* in Acquiring the Second Capsid Shell. J Virol *82*, 11344–11353. 10.1128/JVI.02375-07.

17.   Kato, H., Takahasi, K., and Fujita, T. (2011). RIG-I-like receptors: cytoplasmic sensors for non-self RNA. Immunological Reviews.

18.   Oshiumi, H., Okamoto, M., Fujii, K., Kawanishi, T., Matsumoto, M., Koike, S., and Seya, T. (2011). The TLR3/TICAM-1 Pathway Is Mandatory for Innate Immune Responses to Poliovirus Infection. The Journal of Immunology *187*, 5320–5327. 10.4049/jimmunol.1101503.

19.   Liu, Y., Wang, M., Cheng, A., Yang, Q., Wu, Y., Jia, R., Liu, M., Zhu, D., Chen, S., Zhang, S., et al. (2020). The role of host eIF2α in viral infection. Virol J *17*, 112. 10.1186/s12985-020-01362-6.

20.   Hur, S. (2019). Double-Stranded RNA Sensors and Modulators in Innate Immunity. Annu. Rev. Immunol. *37*, 349–375. 10.1146/annurev-immunol-042718-041356.

21.   Helle Kristiansen, Hans Henrik Gad, Signe Eskildsen-Larsen, Philippe Despres, and Rune Hartmann (2011). The oligoadenylate synthetase family: an

ancient protein family with multiple antiviral activities. J Interferon Cytokine Res. *31*, 41–47. https://doi.org/10.1089/jir.2010.0107.

22.   Han, Y., Donovan, J., Rath, S., Whitney, G., Chitrakar, A., and Korennykh, A. (2014). Structure of Human RNase L Reveals the Basis for Regulated RNA Decay in the IFN Response. *343*.

23.   Bauernfried, S., Scherr, M.J., Pichlmair, A., Duderstadt, K.E., and Hornung, V. (2021). Human NLRP1 is a sensor for double-stranded RNA. INNATE IMMUNITY.

24.   Martinon, F., Burns, K., and Jürg Tschopp The Inflammasome: A Molecular Platform Triggering Activation of Inflammatory Caspases and Processing of proIL-ᴺᴸ. Molecular Cell.

25.   Jin, L., Song, H., Tropea, J.E., Needle, D., Waugh, D.S., Gu, S., and Ji, X. (2019). The molecular mechanism of dsRNA processing by a bacterial Dicer. Nucleic Acids Research *47*, 4707–4720. 10.1093/nar/gkz208.

26.   Lamontagne, B., and Elela, S.A. (2004). Evaluation of the RNA Determinants for Bacterial and Yeast RNase III Binding and Cleavage. Journal of Biological Chemistry *279*, 2231–2241. 10.1074/jbc.M309324200.

27.   Doron, S., Melamed, S., Ofir, G., Leavitt, A., Lopatina, A., Keren, M., Amitai, G., and Sorek, R. (2018). Systematic discovery of antiphage defense systems in the microbial pangenome. Science *359*, eaar4120. 10.1126/science.aar4120.

28.   Periz, J., Celma, C., Jing, B., Pinkney, J.N.M., Roy, P., and Kapanidis, A.N. (2013). Rotavirus mRNAS are released by transcript-specific channels in the double-layered viral capsid. Proc. Natl. Acad. Sci. U.S.A. *110*, 12042–12047. 10.1073/pnas.1220345110.

29.   Ding, K., Celma, C.C., Zhang, X., Chang, T., Shen, W., Atanasov, I., Roy, P., and Zhou, Z.H. (2019). In situ structures of rotavirus polymerase in action and mechanism of mRNA transcription and release. Nat Commun *10*, 2216. 10.1038/s41467-019-10236-7.

30.   McGraw, T., Mindich, L., and Frangione, B. (1986). Nucleotide sequence of the small double-stranded RNA segment of bacteriophage phi 6: novel mechanism of natural translational control. J Virol *58*, 142–151. 10.1128/jvi.58.1.142-151.1986.

31.   Zhang, X., Ding, K., Yu, X., Chang, W., Sun, J., and Hong Zhou, Z. (2015). In situ structures of the segmented genome and RNA polymerase complex inside a dsRNA virus. Nature *527*, 531–534. 10.1038/nature15767.

32.   Ding, K., Nguyen, L., and Zhou, Z.H. (2018). *In Situ* Structures of the Polymerase Complex and RNA Genome Show How Aquareovirus Transcription Machineries Respond to Uncoating. J Virol *92*, e00774-18. 10.1128/JVI.00774-18.

33.   Liu, H., and Cheng, L. (2015). Cryo-EM shows the polymerase structures and a nonspooled genome within a dsRNA virus. Science *349*, 1347–1350. 10.1126/science.aaa4938.

34.   Kumar, D., Yu, X., Crawford, S.E., Moreno, R., Jakana, J., Sankaran, B., Anish, R., Kaundal, S., Hu, L., Estes, M.K., et al. (2020). 2.7 Å cryo-EM structure of rotavirus core protein VP3, a unique capping machine with a helicase activity. Sci. Adv. *6*, eaay6410. 10.1126/sciadv.aay6410.

35.   Lawton, J.A., Estes, M.K., and Prasad, B.V.V. (2001). Identification and Characterization of a Transcription Pause Site in Rotavirus. J Virol *75*, 1632–1642. 10.1128/JVI.75.4.1632-1642.2001.

36.   Borodavka, A., Desselberger, U., and Patton, J.T. (2018). Genome packaging in multi-segmented dsRNA viruses: distinct mechanisms with similar outcomes. Current Opinion in Virology *33*, 106–112. 10.1016/j.coviro.2018.08.001.

37.   Ortega-Esteban, Á., Mata, C.P., Rodríguez-Espinosa, M.J., Luque, D., Irigoyen, N., Rodríguez, J.M., de Pablo, P.J., and Castón, J.R. (2020). Cryo-electron Microscopy Structure, Assembly, and Mechanics Show Morphogenesis and Evolution of Human Picobirnavirus. J Virol *94*. 10.1128/JVI.01542-20.

38.   Duquerroy, S., Da Costa, B., Henry, C., Vigouroux, A., Libersou, S., Lepault, J., Navaza, J., Delmas, B., and Rey, F.A. (2009). The picobirnavirus crystal structure provides functional insights into virion assembly and cell entry. The EMBO Journal *28*, 1655–1665. 10.1038/emboj.2009.109.

39.   Ilca, S.L., Sun, X., El Omari, K., Kotecha, A., de Haas, F., DiMaio, F., Grimes, J.M., Stuart, D.I., Poranen, M.M., and Huiskonen, J.T. (2019). Multiple liquid crystalline geometries of highly compacted nucleic acid in a dsRNA virus. Nature *570*, 252–256. 10.1038/s41586-019-1229-9.

40.   Cui, Y., Zhang, Y., Zhou, K., Sun, J., and Zhou, Z.H. (2019). Conservative transcription in three steps visualized in a double-stranded RNA virus. Nat Struct Mol Biol *26*, 1023–1034. 10.1038/s41594-019-0320-0.

41.   Sun, X., Ilca, S.L., Huiskonen, J.T., and Poranen, M.M. (2018). Dual Role of a Viral Polymerase in Viral Genome Replication and Particle Self-Assembly. mBio *9*, e01242-18. 10.1128/mBio.01242-18.

42.   Sun, G., Zang, Q., Gu, Y., Niu, G., Ding, C., and Zhang, P. (2016). Viral metagenomics analysis of picobirnavirus-positive feces from children with sporadic diarrhea in China. Arch Virol *161*, 971–975. 10.1007/s00705-015-2726-2.

43.   Nemecek, D., Heymann, J.B., Qiao, J., Mindich, L., and Steven, A.C. (2010). Cryo-electron tomography of bacteriophage φ6 procapsids shows random

occupancy of the binding sites for RNA polymerase and packaging NTPase. Journal of Structural Biology *171*, 389–396. 10.1016/j.jsb.2010.06.005.

44.   Sen, A., Heymann, J.B., Cheng, N., Qiao, J., Mindich, L., and Steven, A.C. (2008). Initial Location of the RNA-dependent RNA Polymerase in the Bacteriophage Φ6 Procapsid Determined by Cryo-electron Microscopy. Journal of Biological Chemistry *283*, 12227–12231. 10.1074/jbc.M710508200.

45.   Ilca, S.L., Kotecha, A., Sun, X., Poranen, M.M., Stuart, D.I., and Huiskonen, J.T. (2015). Localized reconstruction of subunits from electron cryomicroscopy images of macromolecular complexes. Nat Commun *6*, 8843. 10.1038/ncomms9843.

46.   Pereira, H.G., Flewett, T.H., Candeias, J.A.N., and Barth, O.M. (1988). A Virus with a Bisegmented Double-stranded RNA Genome in Rat (Oryzomys nigripes) Intestines. Journal of General Virology *69*, 2749–2754. 10.1099/0022-1317-69-11-2749.

47.   Kashnikov, A.Yu., Epifanova, N.V., and Novikova, N.A. (2020). Picobirnaviruses: prevalence, genetic diversity, detection methods. Vestn. VOGiS *24*, 661–672. 10.18699/VJ20.660.

48.   Malik, Y.S., Kumar, N., Sharma, K., Dhama, K., Shabbir, M.Z., Ganesh, B., Kobayashi, N., and Banyai, K. (2014). Epidemiology, Phylogeny, and Evolution of Emerging Enteric *Picobirnaviruses* of Animal Origin and Their Relationship to Human Strains. BioMed Research International *2014*, 1–13. 10.1155/2014/780752.

49.   Gatti, M.S., de Castro, A.F., Ferraz, M.M., Fialho, A.M., and Pereira, H.G. (1989). Viruses with bisegmented double-stranded RNA in pig faeces. Res. Vet. Sci. *47*, 397–398.

50.   van Leeuwen, M., Williams, M.M.W., Koraka, P., Simon, J.H., Smits, S.L., and Osterhaus, A.D.M.E. (2010). Human Picobirnaviruses Identified by Molecular Screening of Diarrhea Samples. J Clin Microbiol *48*, 1787–1794. 10.1128/JCM.02452-09.

51.   Duarte Júnior, J.W.B., Chagas, E.H.N., Serra, A.C.S., Souto, L.C. dos S., da Penha Júnior, E.T., Bandeira, R. da S., e Guimarães, R.J. de P.S., Oliveira, H.G. da S., Sousa, T.K.S., Lopes, C.T. de A., et al. (2021). Ocurrence of rotavirus and picobirnavirus in wild and exotic avian from amazon forest. PLoS Negl Trop Dis *15*, e0008792. 10.1371/journal.pntd.0008792.

52.   Mondal, A., Chakravarti, S., and Majee, S. (2014). Detection of picobirnavirus and rotavirus in diarrhoeic faecal samples of cattle and buffalo calves in Mumbai

metropolis, Western India. Veterinaria Italiana, 357–360. 10.12834/VetIt.1109.10.

53.   Bhattacharya, R., Sahoo, G.C., Nayak, M.K., Rajendran, K., Dutta, P., Mitra, U., Bhattacharya, M.K., Naik, T.N., Bhattacharya, S.K., and Krishnan, T. (2007). Detection of Genogroup I and II human picobirnaviruses showing small genomic RNA profile causing acute watery diarrhoea among children in Kolkata, India. Infection, Genetics and Evolution *7*, 229–238. 10.1016/j.meegid.2006.09.005.

54.   Miguel O. Giordano, Laura C. Martinez, Diego Rinaldi, Susana Guinard, Elizabeth Naretto, Rodolfo Casero, Maria R. Yacci, Ariel R. Depetris, Silvia I. Medeot, and Silvia V. Nates (1998). Detection of Picobirnavirus in HIV-Infected Patients With Diarrhea in Argentina. J Acquir Immune Defic Syndr Hum Retrovirol *18*, 380–383. 10.1097/00042560-199808010-00010.

55.   Giordano, M.O., Martinez, L.C., Rinaldi, D., Espul, C., Martinez, N., Maria B. Isa, Ariel R. Depetris, Silvia I. Medeot, and Silvia V. Nates (1999). Diarrhea and enteric emerging viruses in HIV-infected patients. AIDS Res Hum Retroviruses *15*, 1427–1432. 10.1089/088922299309937.

56.   Rosen, B.I., Fang, Z.-Y., Glass, R.I., and Monroe, S.S. (2000). Cloning of Human Picobirnavirus Genomic Segments and Development of an RT-PCR Detection Assay. Virology *277*, 316–329. 10.1006/viro.2000.0594.

57.   Giordano, M.O., Martinez, L.C., Masachessi, G., Barril, P.A., Ferreyra, L.J., Isa, M.B., Valle, M.C., Massari, P.U., and Nates, S.V. (2011). Evidence of closely related picobirnavirus strains circulating in humans and pigs in Argentina. Journal of Infection *62*, 45–51. 10.1016/j.jinf.2010.09.031.

58.   Bányai, K., Jakab, F., Reuter, G., Bene, J., Uj, M., Melegh, B., and Szücs, G. (2003). Sequence heterogeneity among human picobirnaviruses detected in a gastroenteritis outbreak. Arch. Virol. *148*, 2281–2291. 10.1007/s00705-003-0200-z.

59.   Smits, S.L., Schapendonk, C.M.E., van Beek, J., Vennema, H., Schürch, A.C., Schipper, D., Bodewes, R., Haagmans, B.L., Osterhaus, A.D.M.E., and Koopmans, M.P. (2014). New Viruses in Idiopathic Human Diarrhea Cases, the Netherlands. Emerg. Infect. Dis. *20*. 10.3201/eid2007.140190.

60.   Li, L., Giannitti, F., Ullmann, L.S., Deng, X., Pesavento, P.A., Delwart, E., Pusterla, N., Keyes, C., Low, J., and Aleman, M. (2015). Exploring the virome of diseased horses. Journal of General Virology *96*, 2721–2733. 10.1099/vir.0.000199.

61.   Legoff, J., Resche-Rigon, M., Bouquet, J., Robin, M., Naccache, S.N., Mercier-Delarue, S., Federman, S., Samayoa, E., Rousseau, C., Piron, P., et al. (2017). The

eukaryotic gut virome in hematopoietic stem cell transplantation: new clues in enteric graft-versus-host disease. Nature Medicine *23*, 1080–1085. 10.1038/nm.4380.

62.   Kim, K.W., Allen, D.W., Briese, T., Couper, J.J., Barry, S.C., Colman, P.G., Cotterill, A.M., Davis, E.A., Giles, L.C., Harrison, L.C., et al. (2019). Distinct Gut Virome Profile of Pregnant Women With Type 1 Diabetes in the ENDIA Study. Open Forum Infectious Diseases *6*, ofz025. 10.1093/ofid/ofz025.

63.   Carruyo, G.M., Mateu, G., Martínez, L.C., Pujol, F.H., Nates, S.V., Liprandi, F., and Ludert, J.E. (2008). Molecular Characterization of Porcine Picobirnaviruses and Development of a Specific Reverse Transcription-PCR Assay. J Clin Microbiol *46*, 2402–2405. 10.1128/JCM.00655-08.

64.   Bányai, K., Martella, V., Bogdán, Á., Forgách, P., Jakab, F., Meleg, E., Bíró, H., Melegh, B., and Szűcs, G. (2008). Genogroup I picobirnaviruses in pigs: evidence for genetic diversity and relatedness to human strains. Journal of General Virology *89*, 534–539. 10.1099/vir.0.83134-0.

65.   Ganesh, B., Bányai, K., Kanungo, S., Sur, D., Malik, Y.S., and Kobayashi, N. (2012). Detection and Molecular Characterization of Porcine Picobirnavirus in Feces of Domestic Pigs from Kolkata, India. Indian J Virol *23*, 387–391. 10.1007/s13337-012-0106-z.

66.   Ganesh, B., Banyai, K., Masachessi, G., Mladenova, Z., Nagashima, S., Ghosh, S., Nataraju, S., Pativada, M., Kumar, R., and Kobayashi, N. (2011). Genogroup I picobirnavirus in diarrhoeic foals: Can the horse serve as a natural reservoir for human infection? Vet Res *42*, 52. 10.1186/1297-9716-42-52.

67.   Parrish, C.R., Holmes, E.C., Morens, D.M., Park, E.-C., Burke, D.S., Calisher, C.H., Laughlin, C.A., Saif, L.J., and Daszak, P. (2008). Cross-Species Virus Transmission and the Emergence of New Epidemic Diseases. Microbiol Mol Biol Rev *72*, 457–470. 10.1128/MMBR.00004-08.

68.   Berg, M.G., Forberg, K., Perez, L.J., Luk, K.-C., Meyer, T.V., and Cloherty, G.A. (2021). Emergence of a Distinct Picobirnavirus Genotype Circulating in Patients Hospitalized with Acute Respiratory Illness. Viruses *13*, 2534. 10.3390/v13122534.

69.   Krishnamurthy, S.R., and Wang, D. (2018). Extensive conservation of prokaryotic ribosomal binding sites in known and novel picobirnaviruses. Virology *516*, 108–114. 10.1016/j.virol.2018.01.006.

70.   Boros, Á., Polgár, B., Pankovics, P., Fenyvesi, H., Engelmann, P., Phan, T.G., Delwart, E., and Reuter, G. (2018). Multiple divergent picobirnaviruses with

functional prokaryotic Shine-Dalgarno ribosome binding sites present in cloacal sample of a diarrheic chicken. Virology *525*, 62–72. 10.1016/j.virol.2018.09.008.

71.   Smits, S.L., Poon, L.L.M., van Leeuwen, M., Lau, P.-N., Perera, H.K.K., Peiris, J.S.M., Simon, J.H., and Osterhaus, A.D.M.E. (2011). Genogroup I and II Picobirnaviruses in Respiratory Tracts of Pigs. Emerg. Infect. Dis. *17*, 2328–2330. 10.3201/eid1712.110934.

72.   Huaman, J.L., Pacioni, C., Sarker, S., Doyle, M., Forsyth, D.M., Pople, A., Hampton, J.O., Carvalho, T.G., and Helbig, K.J. (2021). Molecular Epidemiology and Characterization of Picobirnavirus in Wild Deer and Cattle from Australia: Evidence of Genogroup I and II in the Upper Respiratory Tract. Viruses *13*, 1492. 10.3390/v13081492.

73.   Ganesh, B., Nagashima, S., Ghosh, S., Nataraju, S.M., Rajendran, K., Manna, B., Ramamurthy, T., Niyogi, S.K., Kanungo, S., Sur, D., et al. (2011). Detection and molecular characterization of multiple strains of Picobirnavirus causing mixed infection in a diarrhoeic child: Emergence of prototype Genogroup II-like strain in Kolkata, India. Int J Mol Epidemiol Genet *2*, 61–72.

74.   Yinda, C.K., Ghogomu, S.M., Conceição-Neto, N., Beller, L., Deboutte, W., Vanhulle, E., Maes, P., Van Ranst, M., and Matthijnssens, J. (2018). Cameroonian fruit bats harbor divergent viruses, including rotavirus H, bastroviruses, and picobirnaviruses using an alternative genetic code. Virus Evolution *4*. 10.1093/ve/vey008.

75.   Yinda, C.K., Vanhulle, E., Conceição-Neto, N., Beller, L., Deboutte, W., Shi, C., Ghogomu, S.M., Maes, P., Van Ranst, M., and Matthijnssens, J. (2019). Gut Virome Analysis of Cameroonians Reveals High Diversity of Enteric Viruses, Including Potential Interspecies Transmitted Viruses. mSphere *4*, e00585-18. 10.1128/mSphere.00585-18.

76.   Liu, J.-J., Chan, D., Xiang, Y., Williams, H., Li, X.-R., Sniezko, R.A., and Sturrock, R.N. (2016). Characterization of Five Novel Mitoviruses in the White Pine Blister Rust Fungus Cronartium ribicola. PLoS ONE *11*, e0154267. 10.1371/journal.pone.0154267.

77.   Hintz, W.E., Carneiro, J.S., Kassatenko, I., Varga, A., and James, D. (2013). Two novel mitoviruses from a Canadian isolate of the Dutch elm pathogen Ophiostoma novo-ulmi (93–1224). Virol J *10*, 252. 10.1186/1743-422X-10-252.

78.   Kleymann, A., Becker, A.A.M.J., Malik, Y.S., Kobayashi, N., and Ghosh, S. (2020). Detection and Molecular Characterization of Picobirnaviruses (PBVs) in the Mongoose: Identification of a Novel PBV Using an Alternative Genetic Code. Viruses *12*, 99. 10.3390/v12010099.

79.   Ullah, K., Mehmood, A., Chen, X., Dar, M.A., Yang, S., and Zhang, W. (2022). Detection and molecular characterization of picobirnaviruses in the wild birds: Identification of a novel picobirnavirus possessing yeast mitochondrial genetic code. Virus Research *308*, 198624. 10.1016/j.virusres.2021.198624.

80.   Ganesh, B., Masachessi, G., and Mladenova, Z. (2014). Animal Picobirnavirus. Virusdisease *25*, 223–238. 10.1007/s13337-014-0207-y.

81.   Luo, X., Lu, S., Jin, D., Yang, J., Wu, S., and Xu, J. (2018). *Marmota himalayana* in the Qinghai–Tibetan plateau as a special host for bi-segmented and unsegmented picobirnaviruses. Emerging Microbes & Infections *7*, 1–8. 10.1038/s41426-018-0020-6.

82.   Wakuda, M., Pongsuwanna, Y., and Taniguchi, K. (2005). Complete nucleotide sequences of two RNA segments of human picobirnavirus. Journal of Virological Methods *126*, 165–169. 10.1016/j.jviromet.2005.02.010.

83.   Collier, A.M., Lyytinen, O.L., Guo, Y.R., Toh, Y., Poranen, M.M., and Tao, Y.J. (2016). Initiation of RNA Polymerization and Polymerase Encapsidation by a Small dsRNA Virus. PLOS Pathogens *12*, e1005523. 10.1371/journal.ppat.1005523.

84.   Kawakami, E., Watanabe, T., Fujii, K., Goto, H., Watanabe, S., Noda, T., and Kawaoka, Y. (2011). Strand-specific real-time RT-PCR for distinguishing influenza vRNA, cRNA, and mRNA. Journal of Virological Methods *173*, 1–6. 10.1016/j.jviromet.2010.12.014.

85.   Vashist, S., Urena, L., and Goodfellow, I. (2012). Development of a strand specific real-time RT-qPCR assay for the detection and quantitation of murine norovirus RNA. Journal of Virological Methods *184*, 69–76. 10.1016/j.jviromet.2012.05.012.

86.   Brown, T.S., Chadalavada, D.M., and Bevilacqua, P.C. (2004). Design of a Highly Reactive HDV Ribozyme Sequence Uncovers Facilitation of RNA Folding by Alternative Pairings and Physiological Ionic Strength. Journal of Molecular Biology *341*, 695–712. 10.1016/j.jmb.2004.05.071.

87.   Chadalavada, D.M., Cerrone-Szakal, A.L., and Bevilacqua, P.C. (2007). Wild-type is the optimal sequence of the HDV ribozyme under cotranscriptional conditions. RNA *13*, 2189–2201. 10.1261/rna.778107.

88.   Jennifer L. Kopanic, Mona Al-Mugotir, Sydney Zach, Srustidhar Das, Rosslyn Grosely, and Paul L. Sorgen An Escherichia coli strain for expression of the connexin45 carboxyl terminus attached to the 4th transmembrane domain. Frontiers in Pharmacology.

89.   Baca, A.M., and Hol, W.G.J. (2000). Overcoming codon bias: A method for high-level overexpression of Plasmodium and other AT-rich parasite genes in Escherichia coli. International Journal for Parasitology.

90.   Sun, X., Ilca, S.L., Huiskonen, J.T., and Poranen, M.M. (2018). Dual Role of a Viral Polymerase in Viral Genome Replication and Particle Self-Assembly. mBio *9*, e01242-18. 10.1128/mBio.01242-18.

91.   Crosson, S.M., Peter Dib, J. Kennon Smith, and Sergei Zolotukhin (2018). Helper-free Production of Laboratory Grade AAV and Purification by Iodixanol Density Gradient Centrifugation. Mol Ther Methods Clin Dev. *10*, 1–7.

92.   Hodge, K., Tunghirun, C., Kamkaew, M., Limjindaporn, T., Yenchitsomanus, P., and Chimnaronk, S. (2016). Identification of a Conserved RNA-dependent RNA Polymerase (RdRp)-RNA Interface Required for Flaviviral Replication. Journal of Biological Chemistry *291*, 17437–17449. 10.1074/jbc.M116.724013.

93.   Li, X., Gu, M., Zheng, Q., Gao, R., and Liu, X. (2021). Packaging signal of influenza A virus. Virol J *18*, 36. 10.1186/s12985-021-01504-4.

94.   Vázquez, A.L., Alonso, J.M.M., and Parra, F. (2000). Mutation Analysis of the GDD Sequence Motif of a Calicivirus RNA-Dependent RNA Polymerase. J Virol *74*, 3888–3891. 10.1128/JVI.74.8.3888-3891.2000.

95.   Venkataraman, S., Prasad, B., and Selvarajan, R. (2018). RNA Dependent RNA Polymerases: Insights from Structure, Function and Evolution. Viruses *10*, 76. 10.3390/v10020076.

96.   Dominic Mills, J., Kawahara, Y., and Janitz, M. (2013). Strand-Specific RNA-Seq Provides Greater Resolution of Transcriptome Profiling. CG *14*, 173–181. 10.2174/1389202911314030003.

97.   Phan, T., Fay, E.J., Lee, Z., Aron, S., Hu, W.-S., and Langlois, R.A. (2021). Segment-Specific Kinetics of mRNA, cRNA, and vRNA Accumulation during Influenza Virus Infection. J Virol *95*, e02102-20. 10.1128/JVI.02102-20.

98.   Cross, S.T., Michalski, D., Miller, M.R., and Wilusz, J. (2019). RNA regulatory processes in RNA virus biology. WIREs RNA *10*. 10.1002/wrna.1536.

99.   Jaafar, Z.A., and Kieft, J.S. (2019). Viral RNA structure-based strategies to manipulate translation. Nat Rev Microbiol *17*, 110–123. 10.1038/s41579-018-0117-x.

100.   Kobayashi, T., Antar, A.A.R., Boehme, K.W., Danthi, P., Eby, E.A., Guglielmi, K.M., Holm, G.H., Johnson, E.M., Maginnis, M.S., Naik, S., et al. (2007). A Plasmid-Based Reverse Genetics System for Animal Double-Stranded RNA Viruses. Cell Host & Microbe *1*, 147–157. 10.1016/j.chom.2007.03.003.

101.   Kanai, Y., Komoto, S., Kawagishi, T., Nouda, R., Nagasawa, N., Onishi, M., Matsuura, Y., Taniguchi, K., and Kobayashi, T. (2017). Entirely plasmid-based reverse genetics system for rotaviruses. Proc Natl Acad Sci USA *114*, 2349–2354. 10.1073/pnas.1618424114.

102.   Lu, X., McDonald, S.M., Tortorici, M.A., Tao, Y.J., Vasquez-Del Carpio, R., Nibert, M.L., Patton, J.T., and Harrison, S.C. (2008). Mechanism for Coordinated RNA Packaging and Genome Replication by Rotavirus Polymerase VP1. Structure *16*, 1678–1688. 10.1016/j.str.2008.09.006.

103.   Ogden, K.M., Ramanathan, H.N., and Patton, J.T. (2011). Residues of the Rotavirus RNA-Dependent RNA Polymerase Template Entry Tunnel That Mediate RNA Recognition and Genome Replication. J Virol *85*, 1958–1969. 10.1128/JVI.01689-10.

104.   Sun, X., Bamford, D.H., and Poranen, M.M. (2012). Probing, by Self-Assembly, the Number of Potential Binding Sites for Minor Protein Subunits in the Procapsid of Double-Stranded RNA Bacteriophage φ6. J Virol *86*, 12208–12216. 10.1128/JVI.01505-12.

105.   Gottlieb, P., Qiao, X., Strassman, J., Frilander, M., and Mindich, L. (1994). Identification of the Packaging Regions within the Genomic RNA Segments of Bacteriophage Φ6. Virology *200*, 42–47. 10.1006/viro.1994.1160.

106.   Gottlieb, P., and Alimova, A. (2022). RNA Packaging in the Cystovirus Bacteriophages: Dynamic Interactions during Capsid Maturation. IJMS *23*, 2677. 10.3390/ijms23052677.

107.   Mindich, L. (1999). Precise Packaging of the Three Genomic Segments of the Double-Stranded-RNA Bacteriophage φ6. Microbiol Mol Biol Rev *63*, 149–160. 10.1128/MMBR.63.1.149-160.1999.

108.   Thoner, T.W., Meloy, M.M., Long, J.M., Diller, J.R., Slaughter, J.C., and Ogden, K.M. (2022). Reovirus Efficiently Reassorts Genome Segments during Coinfection and Superinfection. Journal of Virology *96*.

109.   Smith, S.C., Gribble, J., Diller, J.R., Wiebe, M.A., Thoner, T.W., Denison, M.R., and Ogden, K.M. (2021). Reovirus RNA Recombination Is Sequence Directed and Generates Internally Deleted Defective Genome Segments during Passage. Journal of Virology *95*.

110.   Fujimura, T., Esteban, R., Esteban, L.M., and Wickner, R.B. (1990). Portable encapsidation signal of the L-A double-stranded RNA virus of S. cerevisiae. Cell *62*, 819–828. 10.1016/0092-8674(90)90125-x.

111.   Chen, S.-C., Born, E. van den, Worm, S.H.E. van den, Pleij, C.W.A., Eric J. Snijder, and René C. L. Olsthoorn (2007). New Structure Model for the Packaging Signal in the Genome of Group IIa Coronaviruses. J. VIROL. *81*, 6771–6774.

112.   Masters, P.S. (2019). Coronavirus genomic RNA packaging.

113.   Narayanan, K., and Makino, S. (2001). Cooperation of an RNA Packaging Signal and a Viral Envelope Protein in Coronavirus RNA Packaging. J Virol *75*, 9059–9067. 10.1128/JVI.75.19.9059-9067.2001.

114.   Kim, D.Y., Firth, A.E., Atasheva, S., Frolova, E.I., and Frolov, I. (2011). Conservation of a Packaging Signal and the Viral Genome RNA Packaging Mechanism in Alphavirus Evolution. J Virol *85*, 8022–8036. 10.1128/JVI.00644-11.

115.   Borodavka, A., Dykeman, E.C., Schrimpf, W., and Lamb, D.C. (2017). Protein-mediated RNA folding governs sequence-specific interactions between rotavirus genome segments. eLife *6*. https://doi.org/10.7554/eLife.27453.001.

116.   Fajardo, T., Sung, P.-Y., and Roy, P. (2015). Disruption of Specific RNA-RNA Interactions in a Double-Stranded RNA Virus Inhibits Genome Packaging and Virus Infectivity. PLOS Pathogens.

117.   Zhou, K., Zhou, L., Lim, Q. 'En, Zou, R., Stephanopoulos, G., and Too, H.-P. (2011). Novel reference genes for quantifying transcriptional responses of Escherichia coli to protein overexpression by quantitative PCR. BMC Molecular Biol *12*, 18. 10.1186/1471-2199-12-18.

118.   Sung, P.-Y., Vaughan, R., Rahman, S.K., Yi, G., Kerviel, A., Kao, C.C., and Roy, P. (2019). The Interaction of Bluetongue Virus VP6 and Genomic RNA Is Essential for Genome Packaging. Journal of Virology *93*, e02023-18. 10.1128/JVI.02023-18.

119.   Da Costa, B., Duquerroy, S., Tarus, B., and Delmas, B. (2011). Picobirnaviruses encode a protein with repeats of the ExxRxNxxxE motif. Virus Research *158*, 251–256. 10.1016/j.virusres.2011.02.018.

120.   Skolnick, J., Gao, M., Zhou, H., and Singh, S. (2021). AlphaFold 2: Why It Works and Its Implications for Understanding the Relationships of Protein Sequence, Structure, and Function. J. Chem. Inf. Model. *61*, 4827–4831. 10.1021/acs.jcim.1c01114.

121.   Jin, S., Contessoto, V.G., Chen, M., Schafer, N.P., Lu, W., Chen, X., Bueno, C., Hajitaheri, A., Sirovetz, B.J., Davtyan, A., et al. (2020). AWSEM-Suite: a protein structure prediction server based on template-guided, coevolutionary-enhanced optimized folding landscapes. Nucleic Acids Research *48*, W25–W30. 10.1093/nar/gkaa356.

122. Mühlmann, M., Eva Forsten, Saskia Noack, and Jochen Büchs (2017). Optimizing recombinant protein expression via automated induction profiling in microtiter plates at different temperatures. Microb Cell Fact *16*.

123. van der Lee, R., Buljan, M., Lang, B., Weatheritt, R.J., Daughdrill, G.W., Dunker, A.K., Fuxreiter, M., Gough, J., Gsponer, J., Jones, D.T., et al. (2014). Classification of Intrinsically Disordered Regions and Proteins. Chem. Rev. *114*, 6589–6631. 10.1021/cr400525m.

124. Mishra, P.M., Navneet Chandra Verma, Chethena Rao, Vladimir N. Uversky, and Chayan Kanti Nandi (2020). Intrinsically disordered proteins of viruses: Involvement in the mechanism of cell regulation and pathogenesis. Progress in Molecular Biology and Translational Science *14*. https://doi.org/10.1016/bs.pmbts.2020.03.001.

125. Hu, G., Katuwawala, A., Wang, K., Wu, Z., Ghadermarzi, S., Gao, J., and Kurgan, L. (2021). flDPnn: Accurate intrinsic disorder prediction with putative propensities of disorder functions. Nat Commun *12*, 4438. 10.1038/s41467-021-24773-7.

126. He, Y., Shivakoti, S., Ding, K., Cui, Y., Roy, P., and Zhou, Z.H. (2019). In situ structures of RNA-dependent RNA polymerase inside bluetongue virus before and after uncoating. Proc. Natl. Acad. Sci. U.S.A. *116*, 16535–16540. 10.1073/pnas.1905849116.

127. Li, X., Zhou, N., Chen, W., Zhu, B., Wang, X., Xu, B., Wang, J., Liu, H., and Cheng, L. (2017). Near-Atomic Resolution Structure Determination of a Cypovirus Capsid and Polymerase Complex Using Cryo-EM at 200kV. Journal of Molecular Biology *429*, 79–87. 10.1016/j.jmb.2016.11.025.

128. Jenni, S., Salgado, E.N., Herrmann, T., Li, Z., Grant, T., Grigorieff, N., Trapani, S., Estrozi, L.F., and Harrison, S.C. (2019). In situ Structure of Rotavirus VP1 RNA-Dependent RNA Polymerase. Journal of Molecular Biology *431*, 3124–3138. 10.1016/j.jmb.2019.06.016.

129. Khlebnikov, A., Risa, Ø., Skaug, T., Carrier, T.A., and Keasling, J.D. (2000). Regulatable Arabinose-Inducible Gene Expression System with Consistent Control in All Cells of a Culture. J. BACTERIOL. *182*.

130. Pradhan, S., Varsani, A., Leff, C., Swanson, C.J., and Hariadi, R.F. (2022). Viral Aggregation: The Knowns and Unknowns.

131. von Martels, J.Z.H., Sadaghian Sadabad, M., Bourgonje, A.R., Blokzijl, T., Dijkstra, G., Faber, K.N., and Harmsen, H.J.M. (2017). The role of gut microbiota in health and disease: In vitro modeling of host-microbe interactions at the aerobe-

anaerobe interphase of the human gut. Anaerobe *44*, 3–12.
10.1016/j.anaerobe.2017.01.001.

132.   Belkaid, Y., and Hand, T.W. (2014). Role of the Microbiota in Immunity and
Inflammation. Cell *157*, 121–141. 10.1016/j.cell.2014.03.011.

133.   Goodacre, N., Devkota, P., Bae, E., Wuchty, S., and Uetz, P. (2020). Protein-
protein interactions of human viruses. Seminars in Cell & Developmental Biology
99, 31–39. 10.1016/j.semcdb.2018.07.018.

134.   Brito, A.F., and Pinney, J.W. (2017). Protein–Protein Interactions in Virus–
Host Systems. Front. Microbiol. *8*, 1557. 10.3389/fmicb.2017.01557.

135.   Kho, Z.Y., and Lal, S.K. (2018). The Human Gut Microbiome – A Potential
Controller of Wellness and Disease. Front. Microbiol. *9*, 1835.
10.3389/fmicb.2018.01835.

136.   Bull, M.J., and Plummer, N.T. (2014). Part 1: The Human Gut Microbiome in
Health and Disease. Integrative Medicine *13*.

137.   Tetter, S., Terasaka, N., Steinauer, A., Bingham, R.J., Clark, S., Andrew J. P.
Scott, Nikesh Patel, Marc Leibundgut, Emma Wroblewski, Nenad Ban, et al.
(2021). Evolution of a virus-like architecture and packaging mechanism in a
repurposed bacterial protein. Science *372*, 1220–1224.

Appendix 1: MS sequencing raw data

Appendix 2: DNA sequence summaries

Gene sequence of hPBV1 viral RNA expression:

T7-hPBV1-<u>HDV</u>

TAATACGACTCACTATAGTAAATTAAATGCTATTTACAAAATTTAAACAGAA
AGGAGAGATGTTATGAACCGGTTTTACAAGGTTTCATACATCGAAAATAGCA
CTACCTGGGGCAGCCGACACACTAACATCGTCTGTTTAACCAGAAGTGTTACT
GAAAGGAGGTTATTTAATGACAGCTAATCAAATTGCCTATCAAAAGCATCTT
GAAACGGCCAGAGTCAATGCCGTGGGTGAGATGCAACGTGGACTTGAACTTG
ATGAGTCCAGGAGGCACAACATTTCTCAGGAACAACTGAAAACACGTGAGTT
AACTGAATTGGAACGCAGCAACCGCGCAGTTGAAAAGGAAACGTCTCGTCAC
AATGTCGTTACCGAAACTGAGACACGCAGATCTAATTTAGCTCGTGAGTGGG
AAACTTATCGTAGCAATTCTGCGCGCGAAATGGAGACTCAGAGGAGTAACAT
TTCCTATGAGGCTATTAAGAGAGGACAGCTAGCTCTGGACCGAGCTGAGCTG
AACGAATCGATACGTGCCACCAACGAAAATCTTGCTTTACAGTATTCTAAATT
ACAAACGGAGAGTTTACTAACTCAACGTGGTCAAGATCTACAACATAAGAAT
GCTATTATAGGTGCAAGCGCCAATGCATTCGGTTCTTTGTTGGGATACTCAAC
AGCAAGTGCGGATCGAGCCAGTCGTGAGGAAATCGCATCCGCTAATCGTAAA
TCACAAGAACATATTGCCAGCATGCAAGTCTTGGGTAGCATGGCCAATACCA
TGTTCTCATCCGTATCCAATCTAGTTGGAAAGACAGCTGGCGCATTTGCAGGA
GGTTTATCATGAAACAGAATGATACTAAGAAAACCACACAACGTCGCAACTC
CAAGAAGTACAGTTCTAAGACGAACCGCGGCACGAAACGTGCGCCGCGCGA
TCAGGAGGTTGGGACGGGTGCGCAAGAAAGTACTCGTAATGACGTTGCTTGG
TATGCTCGTTATCCTCATATTTTGGAGGAAGCTACGCGCCTGCCTTTTGCTTAT
CCTATTGGGCAGTATTATGATACTGGATACTCTGTTGCAAGCGCTACTGAGTG
GTCTAAGTACGTAGATACTAGCCTGACCATTCCTGGTGTGATGTGTGTTAACT
TTACACCCACCCCGGGTGAATCTTATAATAAGAATTCACCTATCAATATAGCT
GCACAGAATGTCTACACGTATGTGCGGCACATGAATTCCGGACACGCCAACT
ATGAGCAGGCTGATCTTATGATGTACTTGCTGGCTATGGACAGTTTGTACATT
TTCCACAGCTATGTTCGGAAAATCCTTGCCATTTCCAAGTTATATACGCCTGT
GAACAAGTATTTCCCGAGAGCTTTGTTAGTAGCTTTGGGTGTTGATCCTGAAG
ATGTTTTCGCTAACCAGGCGCAATGGGAATACTTCGTCAACATGGTGGCATA
CAGGGCTGGAGCGTTCGCTGCTCCTGCAAGTATGACTTATTATGAGCGTCACG
CGTGGATGTCCAATGGTCTGTACGTGGATCAAGATGTCACACGAGCTCAAAT
CTACATGTTCAAGCCCACCATGTTATGGAAGTATGAAAATCTGGGAACCACC
GGTACTAAATTAGTACCACTCATGATGCCCAAAGCTGGAGATAACAGGAAAT
TGGTTGATTTCCAGGTACTGTTCAATAATCTCGTCTCCACTATGTTAGGTGAT
GAAGATTTCGGCATCATGAGCGGTGACGTCTTTAAGGCGTTCGGTGCTGATG
GTCTGGTTAAACTGCTGGCTGTTGACAGCACTACCATGACATTGCCCACATAC
GATCCCCTCATCCTGGCTCAAATTCATAGTGCTAGGGCTGTGGGCGCACCTAT
CCTGGAAACTTCCACTCTTACTGGTTTCCCAGGCCGTCAGTGGCAAATTACAC
AAAACCCTGACGTTAATAATGGCGCCATCATATTCCATCCCTCTTTTGGATAT
GATGGACAGGATCACGAGGAATTATCCTTCCGGGCCATGTGTTCTAACATGA
TTCTCAATCTTCCTGGTGAGGCACACTCGGCGGAGATGATCATTGAGGCTACT

CGCCTGGCTACTATGTTCCAGGTGAAGGCAGTTCCGGCCGGTGATACATCTA
AGCCCGTGCTGTATTTGCCAAATGGTTTCGGGACTGAGGTCGTTAATGACTAC
ACTATGATCAGTGTAGATAAAGCAACTCCACACGACCTCACCATCCATACCT
TCTTCAACAACATTTTGGTTCCTAATGCCAAAGAGAATTATGTTGCGAATCTG
GAACTGCTGAATAACATCATTCAGTTCGATTGGGCTCCTCAGCTCTACCTGAC
GTATGGTATTGCTCAGGAGTCGTTTGGTCCCTTTGCTCAGCTGAATGATTGGA
CCATCCTTACGGGTGAAACTCTGGCTCGGATGCATGAAGTATGTGTCACCAG
CATGTTCGATGTTCCTCAGATGGGCTTCAACAAATAAATGTTGGCCTGGGGTG
TGTGTGCGTTAAGCACACACCTC<u>GGGTCGGCATGGCATCTCCACCTCCTCGCG
GTCCGACCTGGGCATCCGAAGGAGGACGCACGTCCACTCGGATGGCTAAGGG
AG</u>

Gene sequence of ORF2 protein expression:

<u>His</u>-ORF2

ataccATGC<u>ATCATCACCATCACCAC</u>ACAGCTAATCAAATTGCCTATCAAAAGC
ATCTTGAAACGGCCAGAGTCAATGCCGTGGGTGAGATGCAACGTGGACTTGA
ACTTGATGAGTCCAGGAGGCACAACATTTCTCAGGAACAACTGAAAACACGT
GAGTTAACTGAATTGGAACGCAGCAACCGCGCAGTTGAAAAGGAAACGTCTC
GTCACAATGTCGTTACCGAAACTGAGACACGCAGATCTAATTTAGCTCGTGA
GTGGGAAACTTATCGTAGCAATTCTGCGCGCGAAATGGAGACTCAGAGGAGT
AACATTTCCTATGAGGCTATTAAGAGAGGACAGCTAGCTCTGGACCGAGCTG
AGCTGAACGAATCGATACGTGCCACCAACGAAAATCTTGCTTTACAGTATTCT
AAATTACAAACGGAGAGTTTACTAACTCAACGTGGTCAAGATCTACAACATA
AGAATGCTATTATAGGTGCAAGCGCCAATGCATTCGGTTCTTTGTTGGGATAC
TCAACAGCAAGTGCGGATCGAGCCAGTCGTGAGGAAATCGCATCCGCTAATC
GTAAATCACAAGAACATATTGCCAGCATGCAAGTCTTGGGTAGCATGGCCAA
TACCATGTTCTCATCCGTATCCAATCTAGTTGGAAAGACAGCTGGCGCATTTG
CAGGAGGTTTATCATGActcgagc

Gene sequence of hPBV1 viral RNA expression:

T7-hPBV2-<u>HDV</u>

<u>TAATACGACTCACTATA</u>GTAAAATTTTCGAATTTTATAATAATTAAGAAAGG
AGTTTAATAGTTTATCACAACTTAAAAGTGAAATTTATTTAAGAAAGGAGGA
CTACTTATGCAAGTAGCCCCTAATGTATGGAGTAAATATTTTAATATTCCAAA
TCCAGGATTACGTGCTTACTTTAGCAATGTAGTGTCTGGACAGCCGGAGGTAT
ATCGGACGCCATTTTATAAAGGTATGTCTTTGGAATCCATCTGTGACGAGTGG
TACAAGAAACTTGTGTCTATAGACACACAGTGGCCCACCTTGATGGAATTCG
AAGATGACCTGCGGAAGAAGGTCGGTCCGATGTCAGTCATGTTGCCTTTAAA
GGAAAGAATGAGTGACATTGACTCTTACTATGATTCTATCTCCAAGGATCAG
GTTCCCTTCGATACAAAGGCTATTAGCGCGGCCAAATCGGAGTGGAAAGGCG
TATCCCGACTACGCCTGCGCAGCGAGGTTAACACTGTTGCCGTCATGAAGAA
GTCAACCAACAGTGGATCGCCATACTTCTCCAAGCGGAAAGCAGTTGTATCT
AAGACTATACCATGTGATGTGTACATGGATGGTCGATATTGTGTCATGCGCCA
GAATGGTCGTGAATGGTCTGGTGCAGCTGTGCTCGGGTGGCGAGGCCAGGAG

110

GGTGGCCCTAAACCAACAGACGTGAAACAGCGGGTAGTGTGGATGTTTCCTT
TCGCTGTTAATATTCGTGAGCTGCAGGTCTATCAACCACTAATCCTTACGTTC
CAAAGATTGGGATTGGTGCCAGCGTGGGTTAGCATGGAAGCTGTTGATAGGC
GTATCACTAAGATGTTTGATACAAAAGGACCACGTGATGTGGTGGTCTGTAC
AGACTTCTCCAAGTTCGACCAGCATTTTAATCCAACGTGCCAGAGTGTAGCCA
AAGAGCTGTTGGCTGATCTGCTTACAGGTCAGGAAGCCGTCGACTGGTTAGA
ACGTGTGTTCCCAATTAAATATGCTATACCCTTAGCTTATAATTGGGGTGAAA
TCAGGTATGGTATCCATGGAATGGGATCGGGTTCTGGCGGCACTAACGCCGA
TGAGACGTTGGTGCATAGGGTGCTGCAACACGAAGCAGCGATTAGCCACCAT
ACCACTCTTAACCCAAATTCGCAGTGTCTGGGTGACGATGGCGTTTTAACTTA
CCCAGGTATATCTGCGGAGGATGTAATGCAATCATACTCGCGCCACGGTCTC
GATATGAACCTAGAGAAACAGTATGTGAGCAAACAAGACTGCACATATTTAC
GCAGGTGGCATCATACAGATTATCGCGTAGACGGCATGTGTGTGGGAGTGTA
CTCAACCATGCGGGCGTTAGGCAGGTTGGCTATGCAAGAGCGCTACTATGAC
CCAGACGTATGGGGTGAGAAGATGGTCACCCTACGTTATCTATCCATCATTG
AGAATGTGAAGTATCATCCTCTAAAGGAAGAGTTCCTGGACTTTTGCATCAA
AGGGGATAAAACTAGACTTGGACTAGGAATCCCAGGCTTTCTGGACAACATC
GCTGGTGAGGCCCAGAAGGCTATCGACATGATGCCAGATTTCCTTGGCTACA
CCAAGTCGTTGCAGTATGATGGTGATCTGCGACGTAATGCTGCTGCCGGTATT
GAAAACTGGTGGGTTGTTCAAGCATTGAAATCGAGACGCTGACGATCGAGAT
GGTGCAGCAAACCATTGGGACTAACAGTCCCAACTGC<u>GGGTCGGCATGGCAT
CTCCACCTCCTCGCGGTCCGACCTGGGCATCCGAAGGAGGACGCACGTCCAC
TCGGATGGCTAAGGGAG</u>

Appendix 3 CcFV-1 RdRP and MTase

As known in previous chapters, dsRNA viruses infect a broad range of hosts, from bacteria to humans. Most dsRNA viruses encapsidate their dsRNA genomes in the icosahedral capsids to efficiently transcribe and replicate dsRNA inside the virion. The number of packaged RNA segments can vary from 1 to 12. In reoviruses, the RdRP holds the RNA duplex termini and sits at the 5-fold vertices, ready for the transcription upon signals from the capsid conformational changes[29,31,32]. The icosahedral structure of capsids is believed to maximize the efficiency of RdRP transcription and replication. Interestingly, a filamentous dsRNA virus was isolated recently from a tea-infecting fungal pathogen. Though the majority of known polymycoviruses form non-conventional structures, *Colletotrichum camelliae* filamentous virus 1 (CcFV-1) is the only one with a semi-rigid filamentous capsid, which is ~15 nm wide and up to 4,400 nm long[3]. Ten encoded ORFs are predicted on eight RNA segments (Fig. S2). Therefore, the following experiments attempted to unveil why CcFV evolved a filamentous structure and whether its RdRP interacted differently with viral RNA and other proteins inside the filamentous capsids.
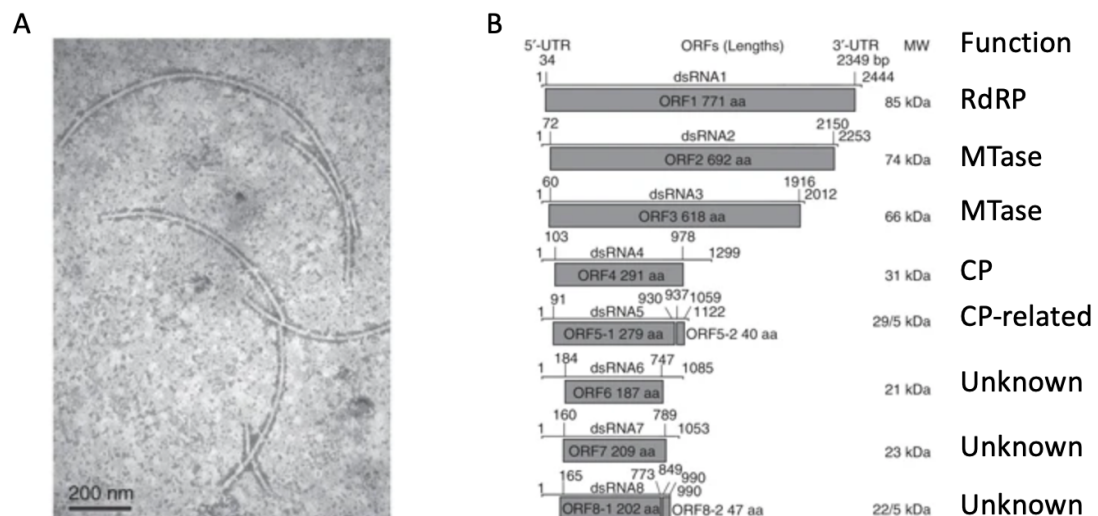

Figure S3: CcFV-1 structure and genome organization. (A) CcFV-1 as filamentous viruses under TEM. (B) The ten predicted ORFs on CcFV-1 eight dsRNA gene segments.

Appendix 3.1. Purification of CcFV-1 RdRP

RdRP structure and replication mechanism were the key to answering the above questions. Therefore, I first attempted to purify CcFV-1 RdRP for structural and biochemical assays. The RdRP was codon-optimized and fused with an N-terminal His-SUMO in pETDuet-His-SUMO-RdRP. RdRP was expressed in *E. coli* Rosetta 2 (DE3) cells by IPTG induction at 16 °C for 20 hours. The cell pellet was lysed in the buffer containing 500 mM NaCl, 50 mM Tris pH 7.5, 10 % glycerol, 15 μg/ml RNase, and 5 mM β-ME. His-SUMO-RdRP protein was purified by Ni-NTA gravity column and eluted with 250 mM imidazole. The eluted His-SUMO-RdRP was incubated with ULP protease at a 10:1 ratio at 4 °C for 16 hours to remove the His-SUMO tag. The ULP cleaved RdRP was eventually loaded to a size-exclusion S200 FPLC to separate the purified RdRP from the cleaved His-SUMO tag. The RdRP dimer was observed in the elution around 62.83 ml (fractions 18-20) from S200 FPLC (Fig. S2.1.1*A*). However, the dimer was not homogenous, with a minor contaminant of a ~ 70 kDa protein observed in
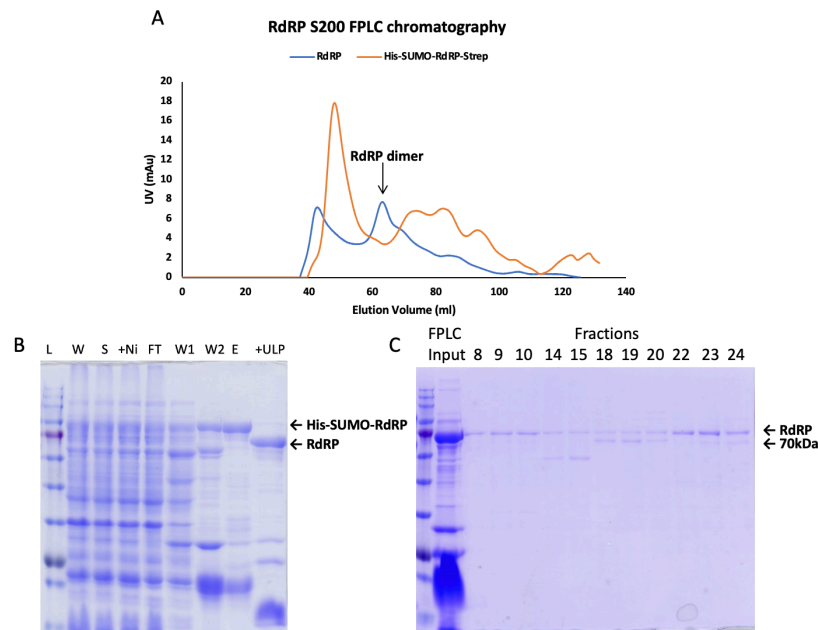

Figure S3.1.1: CcFV-1 RdRP purification. (A) S200 FPLC chromatography of RdRP and His-SUMO-RdRP-Strep. (B) SDS-PAGE of RdRP purification after Ni-NTA gravity column and ULP cleavage. (C) SDS-PAGE of RdRP S200 FPLC fractions.

the same fractions (Fig. S2.1.1*C*). Moreover, the RdRP dimer was unstable, sometimes appearing as monomers under cryoEM. To be noted, the ULP cleaved RdRP was estimated at around 75 kDa, smaller than the expected 85 kDa, indicating possible protein degradation during purification. To optimize the protein purification protocol, His-SUMO-RdRP was purified without ULP digestion or separated by its nucleic acid binding activity through the heparin column before injecting it into the S200 column. Unfortunately, the purified RdRP protein was not separated from the ~70 kDa contaminated protein, and the amount of RdRP dimer was not increased. To characterize the protein degradation during purification, a Strep tag was fused to the His-SUMO-RdRP as His-SUMO-RdRP-Strep. Purified His-SUMO-RdRP-Strep from the Strep column successfully removed the 70 kDa contaminant observed in His-SUMO-RdRP purification. However, His-SUMO-RdRP-Strep eluted only as monomers, which was too small to be solved by cryoEM (Fig. S2.1.1*A*). His-SUMO-RdRP-Strep after ULP cleavage was detected as a ~ 75 kDa protein on SDS-PAGE and anti-Strep Western blots, suggesting an intact C-terminal domain after ULP digestion. Thus, the RdRP degradation occurred in the N-terminal region, supported by the detected peptide results from MS sequencing of the ~75 kDa RdRP after ULP cleavage (Appendix 1.4).

Based on the Alphafold2 structure prediction, the N-terminal domain was flexible, consistent with previous results. Thus, an N-terminal truncated RdRP was constructed with an N-terminal Strep tag (Strep-Δ85). Strep-Δ85 RdRP was eluted in a single peak as monomers (fractions 24-26) from S200 FPLC with a significantly higher yield than the full-length RdRP. The purified Strep-Δ85 RdRP monomers were screened for crystallization using commercial crystal screening trays (NeXtal AmSO$_4$ Suite, NeXtal JCSG+ Suite, PEG/Ion HT, and Wizard 1&2 kits). Unfortunately, no promising crystallization condition was found, with either precipitants or clear drops observed in different buffer conditions. The crystal screening results suggested protein stability or purity issue, which prevented the formation of crystal lattices due to sample heterogeneity. Analyzing the purified monomer by anti-Strep Western blot, with minor protein degradation issues, contaminants were observed at around 48-kDa and 50-kDa without Strep tags. Regarding the observed protein digestion, the purified Strep-Δ85 RdRP was treated by Trypsin or Chymotrypsin digestion for a more stable construct.

Unfortunately, the following smaller stable construct of RdRP was around 45 kDa, which was too small to retain its functional domains for structural studies. The trypsin and chymotrypsin digestion results also suggested that the contaminants at around 48-kDa and 50-kDa in purified Strep-Δ85 RdRP from the S200 column were not byproducts from RdRP protein degradation and thus were contaminants of the host proteins. Further optimization in the purification process to improve the purity of RdRP was required for protein crystallization.
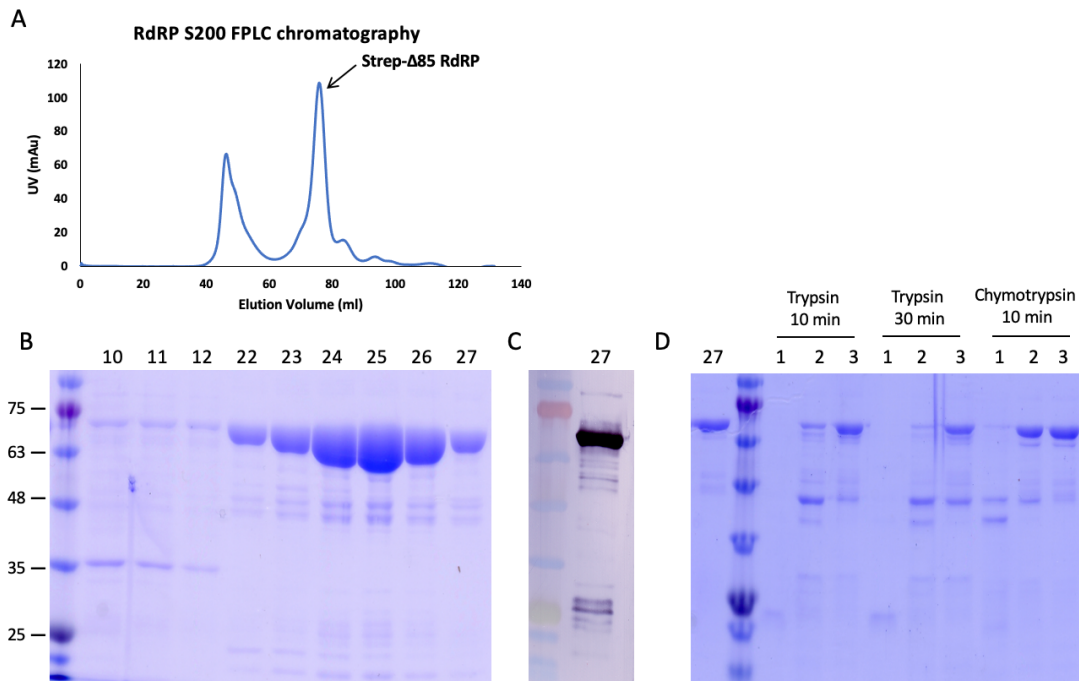

Figure S3.1.2: CcFV-1 Strep-Δ85 RdRP purification. (A) S200 FPLC chromatography of Strep-Δ85 RdRP. (B) SDS-PAGE of Strep-Δ85 RdRP S200 FPLC fractions. Each lane was labeled with the fraction number from S200. (C) Anti-Strep Western blot for Strep-Δ85 RdRP. 27: fraction 27 from S200. (D) Trypsin or chymotrypsin digestion of Strep-Δ85 RdRP. Lane 1: sample mixed in 10:1 protein:protease ratios; Lane 2: sample mixed in 1000:1 protein:protease ratios; Lane 3: sample mixed in 10,000:1 protein:protease ratios.

Appendix 3.2. Purification of RdRP and MTase complex

Since 5'-cap was the key to recruiting ribosomes for protein translation in the eukaryotic system, viral methyltransferase (MTase) or capping enzymes were usually found in eukaryotic dsRNA viruses to cap the nascent RNA transcripts. MTase coupled with RdRP to cap the 5'-end of viral RNA may be either encapsidated or associated with capsids near the packaged RdRP. For example, the turreted reoviruses, *Spinareoviridae*, house their capping enzymes in turrets on the twelve five-fold vertices of the innermost capsid layer, with their RdRPs anchored at the interior of these five-fold vertices[14,31–33]. Non-turreted reoviruses, e.g., rotaviruses, encapsulate their capping enzymes, which are proposed to be close to the packaged RdRP during transcription[34]. To study the interaction between CcFV-1 RdRP and its MTases, RdRP was co-expressed with one of the predicted MTases (ORF3 or P3). The RdRP was expressed from pETDuet-His-SUMO-RdRP, while the MTase P3 from pET28a-Strep-P3-His. The co-expression lysate was incubated with Strep-resins, which could only bind to the N-terminal Strep tag of the MTase P3. As results showed, His-SUMO-RdRP was co-eluted with Strep-P3-His from the Strep column (Fig. S2.2*A*). The identities of RdRP and MTase were confirmed by anti-His and anti-Strep Western blots. The co-eluted RdRP and MTase P3 indicated direct protein-protein interactions between RdRP and MTase P3. To further characterize
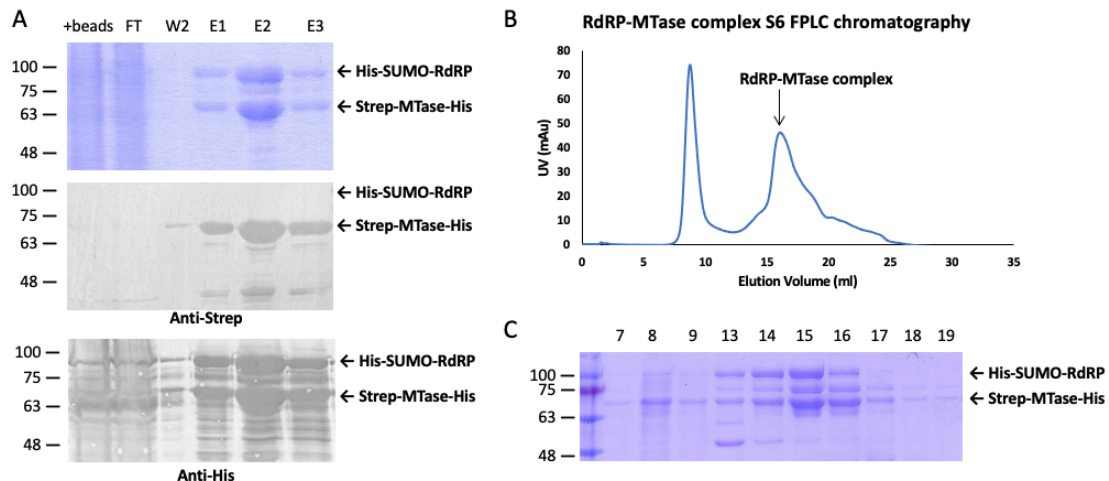

Figure S3.2: CcFV-1 RdRP-MTase complex purification. (A) SDS-PAGE and Western blots for RdRP and MTase co-elution from Strep gravity column. (B) S6 FPLC chromatography of RdRP and MTase complex. (C) SDS-PAGE of RdRP and MTase complex from S6 FPLC.

the co-eluted RdRP-P3 complexes, elution fraction from the Strep column was concentrated and injected into the size-exclusive S6 FPLC. The RdRP-MTase complex was co-eluted as a single peak around 15 ml (fractions 15), with estimated RdRP:MTase ratios as 1 : 1 or 1 : 2 (Fig. S2.*2B-C*).

Future experiments could characterize the interaction of RdRP with the other MTase (ORF2 or P2). Co-expression and pull-down assays of different truncated RdRP mutants and truncated MTase mutants would further illustrate the interaction region between RdRP and the two MTases. The stable RdRP-MTase complexes could also be imaged under cryo-EM for structural studies. In addition, RdRP and MTase enzymatic activity could be characterized by *in vitro* biochemical assays with purified proteins. Structural and biochemical results would eventually reveal the RNA synthesis mechanism of CcFV-1 RdRP.