Determinants of Core Shell Dependent Rotavirus Polymerase Activity

Courtney Long Steger

Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
In
Translational Biology, Medicine, and Health

Sarah M. McDonald, Co-Chair
Pablo Sobrado, Co-Chair
Leslie E. LaConte
Xiang-Jin Meng
Charles J. Schleupner
James W. Smyth

January 31, 2019
Roanoke, Virginia

Keywords: rotavirus, RNA-dependent RNA polymerase, core shell protein, RNA synthesis, genome replication

Determinants of Core Shell Dependent Rotavirus Polymerase Activity

Courtney Long Steger

ACADEMIC ABSTRACT

Rotaviruses (RVs) are medically significant gastrointestinal pathogens and are a leading cause of childhood mortality in many countries. The RV RNA-dependent RNA polymerase, VP1, synthesizes RNA during viral replication only in the presence of another RV protein, VP2, which comprises the innermost core shell layer of the virion. Though these VP1-VP2 interactions are essential for RV replication, the mechanism by which the core shell regulates polymerase activity remains incompletely understood. Here, we sought to identify and characterize specific regions of both VP1 and VP2 that are required for core shell dependent polymerase activity. First, we used bioinformatics approaches to analyze VP1 and VP2 sequence diversity across many RV strains and identify positional locations of critical amino acid changes within the context of known structural domains and motifs. We next tested how the identified sequence differences influenced VP2-dependent VP1 activity *in vitro*. These data revealed that VP1 and VP2 protein diversity correlates with functional differences between avian and mammalian RV strains. Then, we used these sequential and functional incompatibilities to map key regions of VP1 important for mediating RNA synthesis. To pinpoint critical interacting regions of VP1 and VP2, we used site directed mutagenesis to engineer several modified VP1 and VP2 proteins. Then, we employed an *in vitro* RNA synthesis assay to test how the introduced mutations influenced VP2-dependent VP1 activity. Altogether, our results revealed several functionally important VP1 residues critical for *in vitro* VP2-dependent VP1 activity, either individually or in combination with neighboring residues, including E265/L267, R614, and D971/S978/I980. Structural analyses show VP2 interactions at these surface-exposed VP1 sites,

which altogether supports a direct contact model of core shell dependent RV polymerase activity. Moreover, recombinant VP1 proteins containing multiple mutations at buried residues were incapable of facilitating RNA synthesis *in vitro* under the assay conditions, indicating that an extensive intramolecular signaling network exists to mediate VP1 activity. Taken together, these results suggest that VP2 binding at the VP1 surface may induce intramolecular interactions critical for VP1 activity. Overall, results from these studies provide important insight into VP1-VP2 binding interface(s) that are necessary for RV replication.

Determinants of Core Shell Dependent Rotavirus Polymerase Activity

Courtney Long Steger

GENERAL AUDIENCE ABSTRACT

Rotaviruses (RVs) are clinically-significant gastrointestinal pathogens that cause severe diarrhea and dehydration in children. RVs encode a specialized polymerase enzyme, called VP1, which functions to synthesize RNA during viral replication. RNA synthesis activities of VP1 are tightly regulated by another RV protein, VP2, which comprises the innermost core shell layer of the virion. Though these VP1-VP2 interactions are essential for viral replication, the mechanism by which the core shell supports polymerase activity remains poorly understood. Here, we sought to identify and characterize specific regions of both VP1 and VP2 that are essential for polymerase activity in a test tube (i.e., *in vitro*). First, we analyzed VP1 and VP2 sequence diversity across many RV strains. Then, we tested how the identified sequence differences influenced VP2-dependent VP1 activation *in vitro*. To pinpoint critical regions of VP1 and VP2, we next engineered and assayed several mutant proteins. Altogether, our results revealed several functionally important residues of VP1 and VP2, which raises new ideas about VP1-VP2 binding interface(s) that are important for viral replication. Moreover, results from these studies may provide a scientific platform for the rational design of next-generation RV vaccines or antiviral therapeutics.

ACKNOWLEDGEMENTS

This body of work was made possible through countless personal and professional collaborations with many individuals including, but certainly not limited to, those highlighted below. First and foremost, I am eternally grateful for the mentorship given by my advisor, Dr. Sarah McDonald. Her infectious enthusiasm, unwavering support, and passion for science inspire me to be a better, more mindful scientist and person every day. Her guidance and advice have been an invaluable gift that I will treasure forever. I was also incredibly fortunate to learn from many talented scientists in the McDonald Lab, and I am so appreciative of the supportive and intellectually stimulating environments at Virginia Tech Carilion Research Institute and Wake Forest University.

The most heartfelt thanks also go to my amazingly supportive husband, Dan Steger, whose unyielding patience and belief in me motivated me to pursue and persist through the most challenging days. Together, we can achieve anything we put our minds to. Also, to my parents, Howie and Laurie Long, whose confidence and pride in me has never wavered. Thank you for allowing me to be independent, encouraging me to be inquisitive, and motivating me to follow my passion. Special thanks also go to Tim Long, who took a chance and invested in my future at a time when others did not. I am so fortunate to have been supported by an incredible network of family, friends, and peers that have made this journey possible, and this dissertation is dedicated to all of you.

TABLE OF CONENTS

INDEX OF TABLES AND FIGURES

CHAPTER 1:

BACKGROUND AND LITERATURE REVIEW

Section 1.1 Public Health Considerations of Rotavirus

Diarrheal disease is a leading cause of malnutrition and death in children under five years old and causes >525,000 child deaths per year globally, according to latest surveillance data from the World Health Organization (WHO) in 2017 (1-3). Various bacterial, viral, and parasitic organisms including, but not limited to, campylobacter, *Escherichia coli*, rotavirus, norovirus, and *cryptosporidium* can produce symptoms of diarrhea in an infected individual (4). By 2017 WHO estimates, there are approximately 1.7 billion cases of diarrheal disease in children worldwide (4). If left untreated, severe dehydration and fluid loss can lead to death. Fortunately, most diarrheal disease can be effectively treated using rehydration therapies via oral rehydration solution (i.e., a mixture of water, salt, and sugar) or intravenous fluids, zinc supplements, and nutrient-rich diets. However, diarrheal disease remains a significant cause of childhood mortality in developing countries wherein access to adequate healthcare is limited and children who die from diarrhea also suffer from underlying malnutrition (5).

Rotaviruses (RVs) are prominent enteric pathogens that cause acute watery diarrhea in infants and young children around the world and are responsible for an estimated 215,000 deaths in children per year (6). RVs are ubiquitous and highly environmentally stable pathogens; subsequently, RV transmission is difficult to prevent and/or control via adequate sanitation and hygiene (7-9). RVs are shed in the stool of an infected host and are transmitted between hosts primarily via a fecal-oral route (10, 11). One or two days after exposure, RV infection spreads to the small intestine, wherein most RV infection and replication localizes within villus enterocytes

lining the ileum (12). In contrast to pathogenic bacterial-induced diarrhea, RV-induced diarrhea is characterized by relatively little intestinal inflammation, however diarrheal disease manifestation in RV infected hosts is multifactorial, and likely the result of several contributing mechanisms that ultimately disrupt small intestine function and homeostasis (13). For example, malabsorption as a consequence of impaired RV-infected villus enterocytes, activation of the enteric nervous system, and hypersecretion of an RV-induced enterotoxin (NSP4, see Section 1.5G) have all been implicated in diarrheal disease manifestation and severity (12, 14-17). Symptoms of RV-induced diarrheal disease are easily alleviated in individuals living in developed countries with adequate medical infrastructure and access to rehydration therapies, however untreated RV infections in neonates and young children commonly lead to severe dehydration and death in less developed areas of the world.

Several prophylactic RV vaccines are available internationally and four orally-administered, live, attenuated RV vaccines are currently prequalified by the WHO and implemented in national immunization programs in >80 countries (Rotarix™, GSK Biologics; RotaTeq™, Merck and Co; Rotavac™, Bharat Biotech; and RotaSiil™, Serum Institute of India). RV vaccines are primarily aimed to prevent severe gastroenteritis in infants and young children <3 years old because RV disease is most acute and life-threatening within this population. Rotarix™ is a monovalent vaccine containing an attenuated version of the most widely circulated RV strain (G1P[8], see Section 1.2 for details on nomenclature and classification) and is administered to infant children in two doses within the four months of life (18). RotaTeq, however, is a pentavalent vaccine comprised of five human-bovine reassortant virus strains that is administered to infants in a 3-dose series within the first six months of life (19). Both RotaRix™ and RotaTeq™ vaccines are used interchangeably in the United States

and worldwide, and there are no significant differences in efficacy (20). The use of vaccines has yielded significant reductions in RV-related hospitalizations and deaths since their approval and distribution in 2008, with RV childhood mortality rates dropping from ~453,000 deaths/year in 2008 to 215,000 deaths/year in 2013 (6, 21). Although vaccine efficacy ranges from 84-98% in developed countries, significantly reduced vaccine efficacy ranging from 40-60% has been demonstrated in some low-income settings where the vaccines are needed the most (22). Unfortunately, severe disease also occurs at disproportionately high levels in developing countries, with 90% of RV-related deaths occurring in Asian and African countries and >50% of RV-related deaths occurring in just four countries (Democratic Republic of Congo, India, Nigeria, and Pakistan) (6). Although the reasons for these geographic inconsistencies in vaccine efficacy are undoubtedly multifaceted, malnutrition, diverse and variable gut microbiota, co-occurring infections with other pathogens, and unreliable vaccine administration most likely contribute to these outcomes (23). Nevertheless, continued research to increase our understanding of RV replication and pathogenesis is critical to inform the design of improved vaccines.

Section 1.2 History and Classification of Rotavirus

RVs were among the first viruses to be identified as causes of acute viral gastroenteritis in humans, second only to Norwalk virus (now called norovirus) discovered in 1972 (24, 25). After the discovery of RV particles in human duodenal mucosa and feces in 1973, several previously described animal viruses identified in the decade preceding this discovery were also identified as RVs (26). In the last ~50 years, RVs have been identified in >35 different mammalian and avian host species including, but not limited to, humans, pigs, cows, mice,

monkeys, dogs, cats, horses, bats, chickens and ducks. Early electron microscopic images of RV particles revealed that RVs are nonenveloped, triple-shelled particles that encapsidate eleven double-stranded (ds) RNA genome segments within the virion core (27, 28). The distinctive "wheel-like" morphology of RV particles, which forms a smooth outer rim decorated with short spokes, is what inspired the name RV (from the Latin *rota,* meaning "wheel") (29).

RVs are prototypic members of the *Reoviridae* family of viruses in which all family members share morphological properties including one, two, or three concentric, icosahedral protein layers that surround and enclose 9-12 linear dsRNA genome segments. According to the International Committee on Taxonomy of Viruses (ICTV), there are 15 genera delineated within the *Reoviridae* family, including two subfamilies *Spinareovirinae* (classified by the presence of spikes or turrets) and *Sedoreovirinae* (classified by the absence of turrets). The *Rotavirus* genus is included within the *Sedoreovirinae* subfamily, along with the *Orbivirus* genus, which includes notable species like bluetongue virus, which is a triple-shelled, 10-segmented dsRNA virus that infects sheep and other ruminants. *Spinareovirinae* include notable genera such as *Orthoreovirus*, *Aquareovirus*, and *Cypovirus*, which have 10-11 dsRNA genome segments enclosed within 1-2 icosahedral protein layers. Studies using mammalian orthoreovirus, another prototypic species in the *Reoviridae* family, have been particularly useful in inspiring advances in the RV field due to overlaps in morphology, protein function, and viral replication.

The *Rotavirus* genus can be divided into at least eight species, Groups A-H, with tentative Groups I and J (30-32). Group A rotaviruses (RVAs) represent the most well-studied RV species, as they account for the vast majority of disease in humans and many other animals, and are the focus of most of the information detailed in this dissertation, unless otherwise noted. Even within this species, RVAs are incredible genetically diverse, with some viral genome

segments sharing <45% nucleotide sequence identity. Historically, a binary classification system based on antigenic reactivities of two outer capsid proteins: VP7, a glycoprotein comprising the outermost protein layer of the RV virion, and VP4, a protease-sensitive spike protein that decorates the VP7 protein shell (see Section 1.4) (33). Thus, RVA strains were classified into VP7- and VP4-derived serotypes (G-type and P-type, for glycoprotein and protease-sensitive protein, respectively), which is similar to influenza H- and N-type classification systems. In 2008, a more extensive genotyping system was developed for RVAs, wherein each individual genome segment of a given RVA strain is assigned a genotype based upon its nucleotide sequence and percent identity cut-off values (34). In the classification system, each segment is designated by a single letter (defined by the function of the encoded protein) and the assigned genotype is listed as a number based on an established nucleotide percent identity cut-off value. There are currently 20-51 assigned genotypes described for RVA genome segments, summarized in **Fig. 1-1** (35).

Section 1.3 Rotavirus Genome Structure and Organization

The RV genome totals ~23 kilobases (kb) and is comprised of eleven linear dsRNA genome segments, numbered (i.e., g1-g11) according to their migration in polyacrylamide gels (**Fig. 1-1**). While the genome segments range in length (0.5-3.3 kb) and composition, they are conserved in their structure and organization. The intra-particle structural organization of the genome is highly ordered and appears to be conserved among other *Reoviridae* viruses including aquareovirus, cypovirus, orthoreovirus, and bluetongue virus (36-39). Structural data of intact virion particles suggest that the dsRNA genome segments are tightly packed and maintain rigid, rod-like morphologies that extend tangentially from each polymerase complex and form parallel

striations (36, 38). Connections between multiple concentric layers of genomic RNA within the core may indicate the presence of complex inter-segment interactions that contribute to, or are the outcome of, whole genome organization (36, 38).

The positive-sense (+) RNA strand of all eleven RV genome segments (+RNA, see Section 1.5C) are arranged with a central coding region in the open reading frame (ORF), which is flanked by 5' and 3' noncoding untranslated regions (UTRs) of varying lengths (5' UTR ranges from 9-48 nucleotides; 3' UTR ranges from 17-182 nucleotides) (40). Noncoding regions vary between RV genome segments, but are highly conserved between homologous strains (41) and contain terminal consensus sequences that are important signals for gene expression and genome replication (42-44). Each +RNA is monocistronic, with the possible exception of g11,

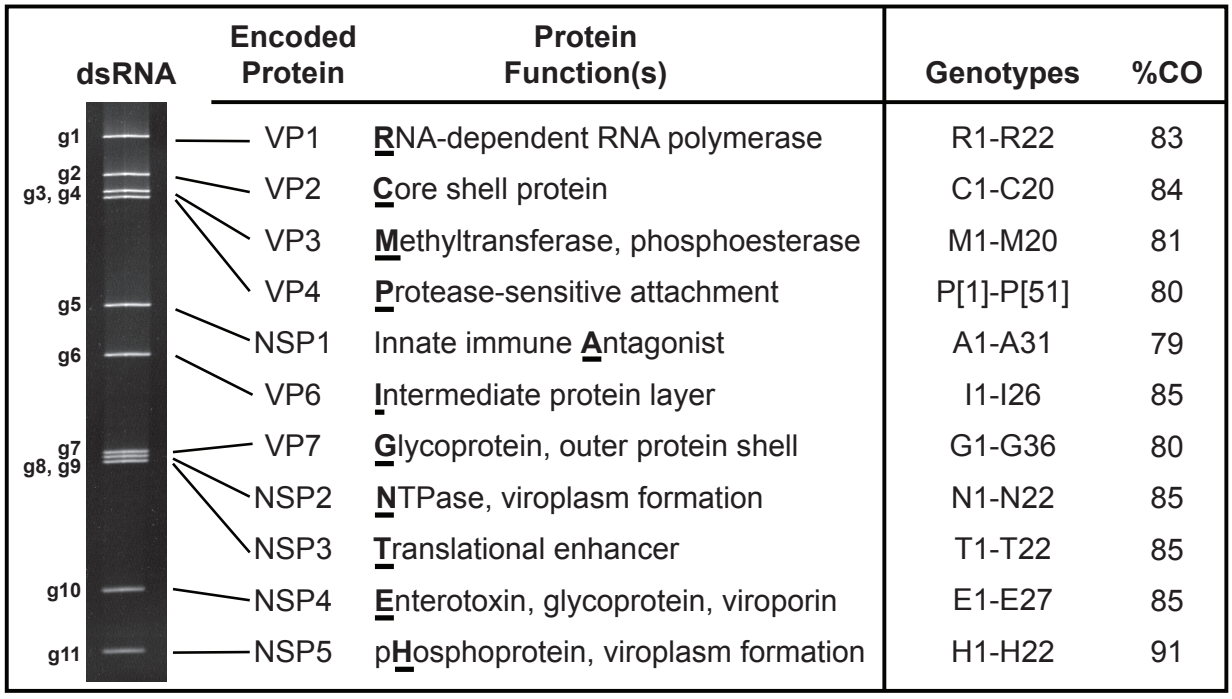


Figure 1-1. Rotavirus gene segments, protein functions, and classification nomenclature. Eleven dsRNA gene segments (rotavirus strain SA11) are shown electrophoresed on a 1% agarose gel with encoded proteins and key protein function(s) labeled. Genetype acronyms are derived from protein function and highlighted in underlined bold font. Genotype numbers are assigned based on percent nucleotide identity cut-off values (%CO), whereby genes that share < %CO value are assigned to a novel genotype.

which encodes for a second protein in some select RV strains (45). Thus, each RV genome segment is considered to encode one viral gene and one viral protein. The eleven distinct RV genome segments are encased within the core of the virion and encode for 6 structural proteins (VP1-VP4, VP6 and VP7) that comprise the infectious RV particle structure, as well as 5 or 6 non-structural proteins (NSP1-NSP6) that have various functions in viral replication, assembly of new virions, and evasion of host antiviral responses (**Fig. 1-1**).

Section 1.4 Rotavirus Particle Structure

The fully infectious RV virion is a non-enveloped, triple-layered particle (TLP) that is ~100 nm in diameter (**Fig. 1-2**). Many structural studies of RV particles have revealed detailed images of particle morphology (46-51). Each concentric protein layer of the RV virion occupies icosahedral symmetry; the outer and middle protein layers possess T=13 icosahedral surface lattice, whereas the innermost core shell layer is organized as a T=1 icosahedron. Near-atomic resolution structures of the TLP also revealed 132 large aqueous channels that penetrate the inner core: 12 Type I channels located at the icosahedral fivefold axes and two sets of 60 Type II or Type III channels located at twofold and threefold axes, respectively (52). Since RV particles remain intact during +RNA synthesis, it is likely that these channels allow for nucleotide exchange and extrusion of nascent transcripts (see Section 1.5C).

The smooth outermost shell of the TLP is comprised of 260 homotrimers of the glycoprotein VP7 (37 kDa monomer) embedded with 60 trimeric VP4 spikes positioned at the fivefold icosahedral axes. The VP4 spikes radially project ~120 Å from the surface of the VP7 shell and also extend inwards ~80 Å to be stabilized by interactions with both the outer and intermediate layers of the TLP, formed by trimers of VP7 or VP6, respectively (50). Exposure

of the TLP to proteases in the intestinal lumen (or trypsin in cell culture), cleaves VP4 (88 kDa) into fragments designated VP8* (28 kDa) and VP5* (60 kDa), which facilitates viral membrane penetration (53). The VP7 protein layer also undergoes significant conformational changes following endocytosis of the RV virion, whereby calcium-dependent VP7 trimer structures are disassembled and the outermost layer of the TLP is shed (see Section 1.5B). VP7 glycoproteins and VP4 spikes, specifically VP5* fragments, are targets for neutralizing antibodies, and thus are important mediators of RV infectivity (see section 1.6B) (54).

The outer VP7 shell directly encloses an intermediate protein layer comprised of 260 homotrimers of VP6 (45 kDa). At 70 nm in diameter, the VP6 protein shell is >20 times thicker that either the outermost VP7 layer or innermost VP2 layer and encompasses the majority of the RV virion. Thus, the VP6 protein layer is important for stabilization of the fragile core particle and maintaining the integrity of the virion structure (46, 49, 51). Additionally, maintaining the

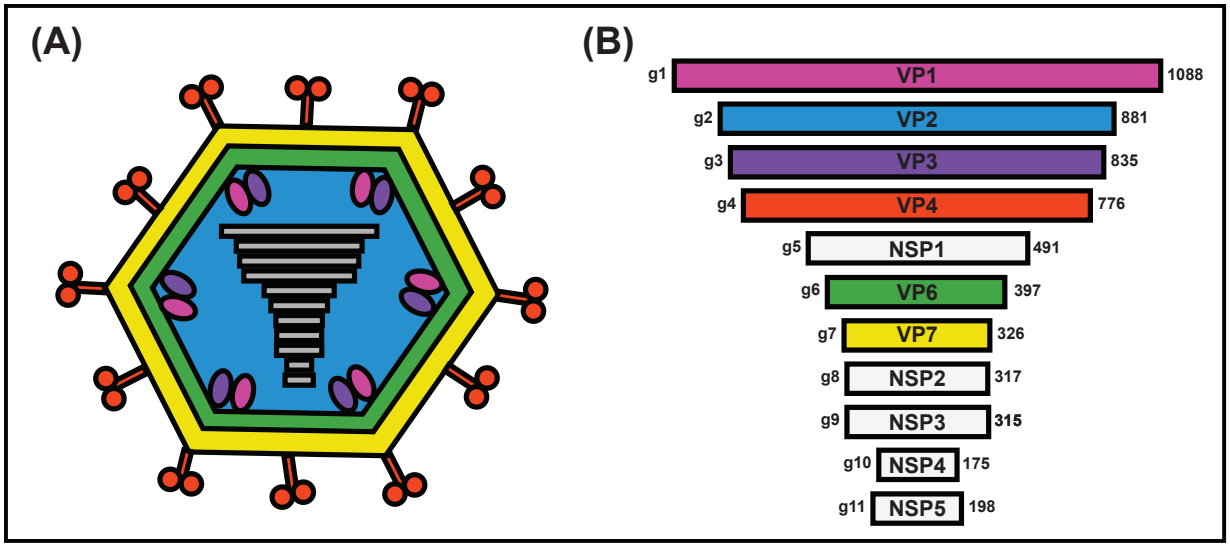


Figure 1-2. Schematic diagrams of rotavirus virion and genome structure. (A) The infectious rotavirus virion is composed of VP4 (red), VP7 (yellow), VP6 (green), VP2 (blue), VP3 (purple), VP1 (pink) and genomic dsRNA (gray). (B) Encoded proteins (in amino acids) of the dsRNA gene segments, with structural proteins colored as in panel A and nonstructural proteins shown in white.

structural integrity of the uncoated TLP (i.e., double-layered particle, or DLP) during transcription is a critical feature of the intermediate protein shell.

Directly beneath the VP6 layer resides the core shell layer, comprised of 120 copies of the VP2 monomer (102 kDa). In order to achieve T=1 icosahedral symmetry, VP2 molecules arrange as dimers, wherein one structural isoform converges tightly around fivefold vertices (VP2-A), and molecules adopting a slightly different conformation (VP2-B) intercalate between adjacent VP2-A monomers to form flower-like decameric structures surrounding each fivefold icosahedral axis (46, 47, 49, 51). The VP2 monomer is comprised of two domains: a principal scaffold domain that creates the thin, comma-like, core shell, and an N terminal domain extension (~100 residues) (49). The principal scaffold domain of VP2 maintains extensive interactions between structural isoforms and is organized into three subdomains (apical, carapace, and dimer-forming) (49). The VP2 NTD forms a flexible 'arm' that is not fully resolved in any known structures, but presumably extends down into the core. Given the non-specific RNA-binding affinity and dynamic mobility of the VP2 NTD, it is likely that the VP2 NTD has a role in RNA synthesis (e.g., transcription and/or genome replication), particle assembly, and encapsidation of viral RNA or proteins. Eleven copies of the viral polymerase (VP1, 125 kDa) and capping enzyme (VP3, 98 kDa) are enclosed within the VP2 core shell layer, along with the dsRNA genome. VP1s are positioned slightly off-center from the fivefold axes (i.e., beneath Type I channels), tethered to the VP2 core shell (55). Atomic resolution structures for a recombinant VP1 molecule revealed the polymerase structure in fine detail (see Section 1.5F) (43), whereas the structure of VP3, which assumedly located adjacent to VP1, remains mostly unsolved (56-58).

Section 1.5 Rotavirus Replication Cycle

1.5A. Attachment of the Rotavirus Particle. RV attachment to intestinal epithelial cells is a complex and incompletely understood area of RV biology. Available data suggests that attachment of the RV particle to host cell receptors is broadly mediated by initial binding to a common glycan receptor, followed by more specific interactions with co-receptors that facilitate virus entry (59). Both VP7 and VP4 are clear drivers of attachment through initial contacts with host cells, however, the precise interactions and mechanistic insights of attachment and penetration are still active areas of investigation. Following protease-mediated VP4 cleavage events, sialic acid-binding residues are exposed on the surface of VP8* fragments noncovalently bound to the TLP (60). Early work indicated that VP8 interactions with sialic acid residues was the key determinant for RV entry and infectivity (61-63). But more extensive analysis demonstrated that although some animal RV strains require sialic acid for attachment, most human and animal strains do not hemagglutinate red blood cells and are neuraminidase-resistant; thus, most RV strains tested to date do not require sialic acid residues for host cell entry (64-66). More recent data has shown that some sialic acid-independent VP8* fragments bind to nonsialylated histo-blood group antigens, which are common cellular receptors for other gastrointestinal pathogens including noroviruses (67). Still, there is considerable strain-specificity in glycan binding, which may have important implications for interspecies transmission, strain susceptibility, and vaccine response (68).

In addition to cellular glycan interactions, coordinated interactions with multiple co-receptors on the cell surface have been proposed to facilitate virus attachment. For instance, VP5* and VP7 glycoproteins are involved in putative binding interactions with various integrins post-attachment (e.g., $\alpha 2\beta 1$, $\alpha v\beta 3$, $\alpha x\beta 2$, $\alpha 4\beta 1$) (69-72). Heat shock cognate protein 70 (hsc70)

and the tight junction protein JAM-A are also implicated as possible RV co-receptors in some strains (73, 74), however the order of interaction(s) is not yet known. It also remains to be determined whether some or all of the implicated molecules work in tandem or represent alternative binding footprints. Knowledge of these co-receptor interactions is limited by a lack of structural data for protein-receptor complexes and unknown binding affinities for receptor molecules. However, all known co-receptors are associated with detergent-resistant lipid rafts, which are required to facilitate efficient interactions between the RV particle and host cell receptor(s) (65, 66).

Interactions between the RV particle and cellular receptor(s) are extensive and indisputably complex. In fact, inhibiting or silencing the interaction of RV with each of the proposed receptors or co-receptors reduces RV infectivity by less than one log, suggesting that there may be additional entry factors that have not yet been described, and/or that RVs have evolved to accommodate more than one entry mechanism (75). Additionally, studies analyzing different receptors using different RV strains on different cell types have yielded conflicting results that are difficult to interpret. A major limitation of many of these studies using non-polarized, lab-adapted cell types (e.g., MA104 cells) is that the complex morphologies characteristic of the natural RV host cells are not fully captured. Future studies will benefit from a human intestinal organoid system that recapitulates the microenvironment of the small intestine in a biologically-relevant *ex vivo* system (76, 77).

1.5B. Entry and Uncoating of the RV Virion. Following attachment to a susceptible intestinal epithelial cell, multiple coordinated processes facilitate entry of a RV virion entry into the host cell. Recent studies indicate that, similarly to attachment mechanisms, the mode by which RVs enter the cell may differ based on RV strain (78). Though the mechanism of

penetration is largely unknown, efficient RV entry is dependent upon post-attachment conformational changes of VP5*. Specifically, the VP5* cleavage fragment undergoes structural rearrangements to expose lipophilic domains of VP5* that facilitate membrane penetration (79). Once the RV virion is internalized, diverse endocytic pathways have been suggested as mechanisms of RV entry, though most human and animal RV strains enter cells via clathrin-mediated endocytosis (78, 80). Unsurprisingly, strain-specific variation in the VP4 spike protein regulates the method by which RV enters the cell (e.g., receptor-mediated endocytosis or clathrin- or calveolin-independent endocytosis). Regardless of entry method, RV particles within calcium-deficient endosomal vesicles undergo a second morphological change wherein the outermost VP7 shell dissociates and solubilizes (81, 82). The precise mechanism of endosomal membrane penetration remains unclear, though initial data suggests that solubilized VP7 proteins can permeabilize cellular membranes and lyse the endosomal vesicle (82) and more recent structural data showing membrane disruption and gross morphological changes supports this model (83). Ultimately, a transcriptionally-active uncoated virion (i.e., DLP) is deposited into the cytosol.

1.5C. Transcription and Translation of Viral +RNAs. Transcription is triggered soon after the DLP is released from the endosome and delivered into the cytosol. Although transcriptional activation is associated with detachment of the outer virion layer, which allows the DLP to expand, it remains mechanistically unclear how uncoating of the TLP induces transcription and if the increase in particle diameter is sufficient to initiate RNA synthesis (48, 84). Nevertheless, within the confines of the fully-formed DLP, each dsRNA genome segment associates with an endogenous RNA-dependent RNA polymerase (VP1) and a viral RNA capping enzyme (VP3), both of which are presumed to be anchored against the innermost core

shell layer beneath each fivefold icosahedral axis. VP1 synthesizes viral +RNAs that are subsequently capped with an m^7GpppG structure on the 5' end via activities of VP3. It is not currently known if the polymerases transcribe dsRNA genome segments simultaneously, or if they function independently, though evidence suggests that transcription of the RV genome is asymmetric in that the amount of +RNAs produced from the dsRNA genome is not equimolar (85, 86). Regardless, newly-synthesized +RNAs are extruded from the core via Type I channels that permeate the VP2 and VP6 layers at fivefold vertices (46, 87). Noncoding regions at the 5' and 3' termini of the +RNAs are partially-inverted complements that confer long-range interactions and create modified panhandle structures (88). Each RV gene lacks 3' polyadenylation signals, but instead contain a highly conserved, single-stranded consensus sequence (5'-UGUGACC-3') that contains important signals for gene expression and genome replication (42-44).

Nascent capped +RNAs released into the cytosol serve as templates for translation by host cell translation machinery, or as templates for genome replication (see Section 1.5F) (**Fig. 1-3**). Both structural (VP) and nonstructural (NSP) RV proteins are synthesized on free ribosomes in the cytosol. Additionally, RV glycoproteins VP7 and NSP4 are synthesized by host cell ribosomes bound to the endoplasmic reticulum (ER), then co-translationally inserted into the ER membrane, during a process that aids in virion maturation and egress (see Section 1.5G). Although viral proteins are synthesized by cellular translation machinery, translation of non-polyadenylated viral +RNAs is facilitated by actions of NSP3 (36 kDa), a functional homolog of cellular poly(A)-binding protein (PABP). More specifically, the NSP3 N-terminus recognizes the highly conserved 3' consensus sequence (3'-GACC) characteristic of all RV +RNAs, and the NSP3 C-terminus binds eIF4G (eukaryotic translation initiation factor 4G) with higher affinity

than PABP to induce translation (89, 90). Thus, competitive binding of NSP3 to eIF4G simultaneously inhibits host cell translation and enhances translation of viral proteins. Additionally, the ssRNA binding capacity of NSP3 has also been proposed as a possible

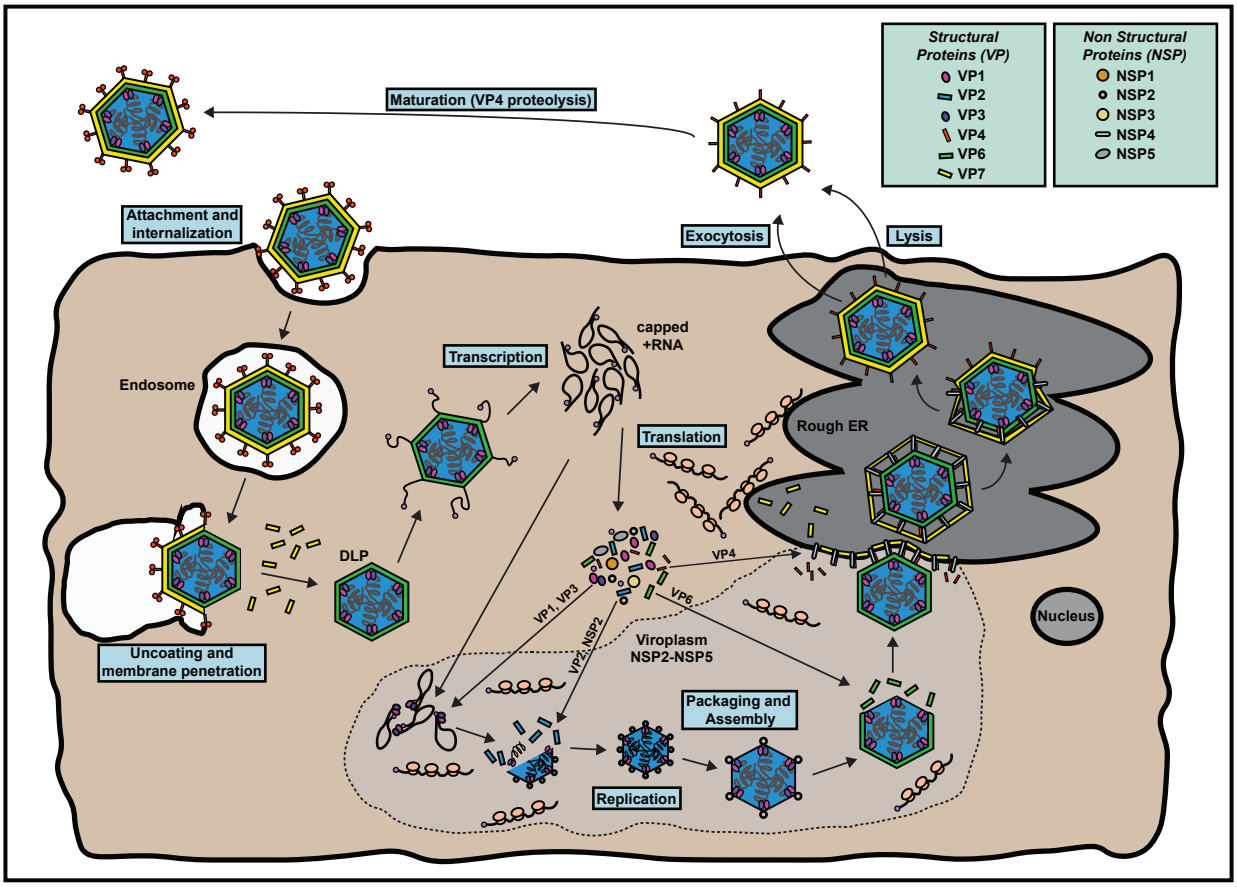


Figure 1-3. Overview of the intracellular rotavirus lifecycle. Fully infectious rotavirus virions enter an infected cell via endocytic processes following attachment of VP4 (red) cleavage products to various surface receptors. Low calcium concentrations within the endosome trigger the triple layered rotavirus particle to shed its outermost layer of VP7 (yellow) and VP4. The double layered particle is deposited into the cytosol and VP1s (pink) contained within the core transcribe the dsRNA genome. Viral +RNA transcripts are capped by the activities of VP3 (purple) and subsequently used as templates for viral protein synthesis in the cytosol or the transcripts are trafficked to the viroplasm (formed by NSP2 and NSP5) and serve as templates for dsRNA synthesis. +RNA templates are packaged within concurrently assembling viral cores and replication of each gene by VP1s occurs in a coordinated and synchronized mechanism within a subviral assembly-replicase intermediate particle that includes VP2 (blue) and possibly VP6 (green) protein layers. Through interactions with NSP4 and VP4, the newly-formed double layered particle buds through the ER membrane and acquires a transient envelope that is displaced by the ER-resident VP7 glycoprotein. Finally, the triple-layered progeny virion exits the cell via lysis or exocytosis.

mechanism to traffic newly synthesized viral +RNAs to viral factories called viroplasms, wherein +RNAs are assorted and packaged with newly forming RV particles and genome replication occurs (91, 92).

1.5D Viroplasm Formation. Virus-induced inclusion bodies known as viroplasms are formed in a concentration-dependent manner regulated by accumulation of two virally-encoded proteins: NSP2 (35 kDa) and NSP5 (26-35 kDa). Although several other viral protein entities localize to these sites (including VP1, VP2, VP3, and VP6), abundant expression of NSP2 and NSP5 only are sufficient to induce self-assembly of viroplasms within an RV infected cell (93). Cellular components including microtubules, lipid complexes, and proteasomes have also been implicated in viroplasm formation (94-96). Altogether, these data suggest that formation of the viroplasm concurrently enhances viral replication and inhibits cellular mechanisms as a consequence of hijacking the cytoskeletal infrastructure. In addition to usurping cellular pathways to promote RV replication, viroplasms function to sequester viral RNAs to stimulate assembly (97). Viroplasmic functions during genome replication are largely mediated by RNA-protein and protein-protein interactions involving NSP2 and NSP5. NSP2, which is the most abundant protein species found in the viroplasm, is a multifunctional protein with enzymatic activities (e.g., NTPase, NDP kinase, RTPase, and helicase activities) and ssRNA-binding capabilities (98-101). Doughnut-shaped, octameric formations of NSP2 create charged grooves that critically engage ssRNA and other proteins including, but not limited to, VP1, VP2, NSP5 and tubulin (102-104). In fact, NSP5 and ssRNAs competitively bind the same site, which may represent a mechanism of functional regulation during genome replication and virus assembly (102).

1.5E +RNA Assortment and Early Particle Assembly. RV +RNA assortment and encapsidation, particle assembly, and genome replication are highly coordinated processes that remain poorly understood due to limited *in vitro* assays and only recent availability of a reverse genetics system. Nevertheless, the available data suggests a model in which *cis*-acting elements between +RNAs encoding the eleven distinct RV genome segments engage each other to form a supramolecular complex, which subsequently gets encapsidated via interactions with core viral proteins (105, 106). More recent *in vitro* data has supported this 'supramolecular' model and moreover identified sequence-specific inter-segment interactions between RV +RNAs that are mediated by structural changes induced by NSP2 binding (107). Packaging signals that facilitate encapsidation of these +RNA interaction networks into the viral core are remain to be determined, though segment-specific packaging signals are predicted to reside, at least in part, within 5' and 3' UTR flanking the ORF of each +RNA. Sequences within these terminal regions differ for each of the eleven +RNAs, but are highly conserved among homologous gene segments from various RV strains and likely contain structural elements involved in assortment and packaging (47, 108).

Though the RV packaging efficiency has not been experimentally calculated, all naturally-occurring RV strains that have been isolated to date contain eleven genome segments that are present at equimolar amounts at the population level (109). Moreover, there have been no reports of strains that lack one or more gene segment(s), despite the nonessential role of NSP1 in the RV replication cycle in culture (97, 110-112). Altogether these data provide convincing evidence that there are strong RNA-RNA and/or RNA-protein interactions among the +RNA segments that contribute to an all-or-nothing packaging system wherein a total of eleven genome segments (i.e., one copy of each gene) are invariably packaged within an assembling RV particle

(107, 109, 113 2015). In this model, encapsidation of a supramolecular RNA complex containing all eleven +RNA gene segments is required for genome replication to take place. Once packaged within a nascent viral core particle, each +RNA associates with a distinct polymerase complex that functions in concert with ten other polymerase complexes to form the complete dsRNA genome.

Coordinated synthesis of eleven dsRNA segments occurs within the confines of a subviral assembly-replicase intermediate (RI) particle, presumably to sequester the rigid dsRNAs within a newly forming virion (114). Although the order of assembly and replication is still debated, several distinct RI particles have been identified and characterized. Coupled assembly-replication processes have been difficult to study in the context of infected cells because they occur within electron-dense viroplasmic inclusion bodies that cannot be resolved using transmission electron microscopy (EM). However, subcellular fractions containing RV RIs capable of mediating *in vitro* dsRNA synthesis have been isolated biochemically and visualized by EM, which revealed a heterogeneous population of particles ranging in size (30-nm to 80-nm) and morphology (smooth versus rough borders) (87, 106, 115, 116). Altogether, these data suggest a stepwise model of assembly (shown in **Fig. 1-3**) wherein small, smooth particles representing viral cores (i.e., +RNAs bound to VP1/VP3 polymerase complex within a partially-formed VP2 capsid layer) acquire an incomplete VP6 layer visually represented by larger, rough particles that display a honeycomb surface (114, 116). A more detailed review of these processes and the proposed model of RV particle assembly can be found in Chapter 2 of this dissertation.

1.5F Synthesis of dsRNA Genome Segments. Replication of the dsRNA genome is precisely regulated by complex interactions that are incompletely understood. However, insight

into how VP1 mediates both transcription and genome replication have been gleaned from high-resolution structural analysis of the recombinant (r) VP1 structure in addition to structures of the RV DLP. VP1 assumes a compact and globular morphology that creates a cage-like structure with a buried catalytic core (43). The tri-domain organization of VP1 contributes to the formation of four distinct tunnels that permeate the buried catalytic core and allow for separate entry and exit of +RNA templates, extrusion of nascent dsRNA genome segments, and exchange of essential nucleotides and metal ions (43). Analyses of DLP structures suggest that VP1 is positioned directly beneath the core shell such that the +RNA exit tunnel is oriented against VP2, which creates several overlapping regions at this interface (55). Such an orientation would facilitate extrusion of newly-formed +RNAs out of the DLP via adjacent Type I channels at the fivefold vertices and is similar to proposed orientations of polymerase complexes of other *Reoviridae* members (37, 38, 117). However, the functional significance of these interaction interfaces in the context of transcription have yet to be fully explored, though we provide insight into the importance of these overlapping residues during genome replication within Chapter 4 of this dissertation.

Interestingly, VP2 is required for polymerase activation during both transcription and genome replication (42, 118-120). In fact, the crystal structure of rVP1 in complex with +RNA template revealed an auto-inhibited, inactive polymerase (43). Comparisons to the enzymatically active orthoreovirus polymerase crystal structure distinguished several structural elements including a misaligned +RNA template and a malpositioned priming loop that inhibit initiation (**Fig 1-4**) (43). A leading hypothesis in the field posits that VP2 binding to VP1 may trigger a signaling cascade within the VP1 enzyme that resolves the aforementioned structural anomalies and ultimately triggers polymerase activation. Biochemical details of VP1 activation during

genome replication have been resolved using *in vitro* replication systems (118, 121). Early studies using a cell-free open core replication system and an *in vitro* RNA synthesis assay revealed that dsRNA synthesis is +RNA template-dependent and also requires GTP and divalent metal cation co-factors in addition to the VP2 activator protein (42, 122-124). More specifically,

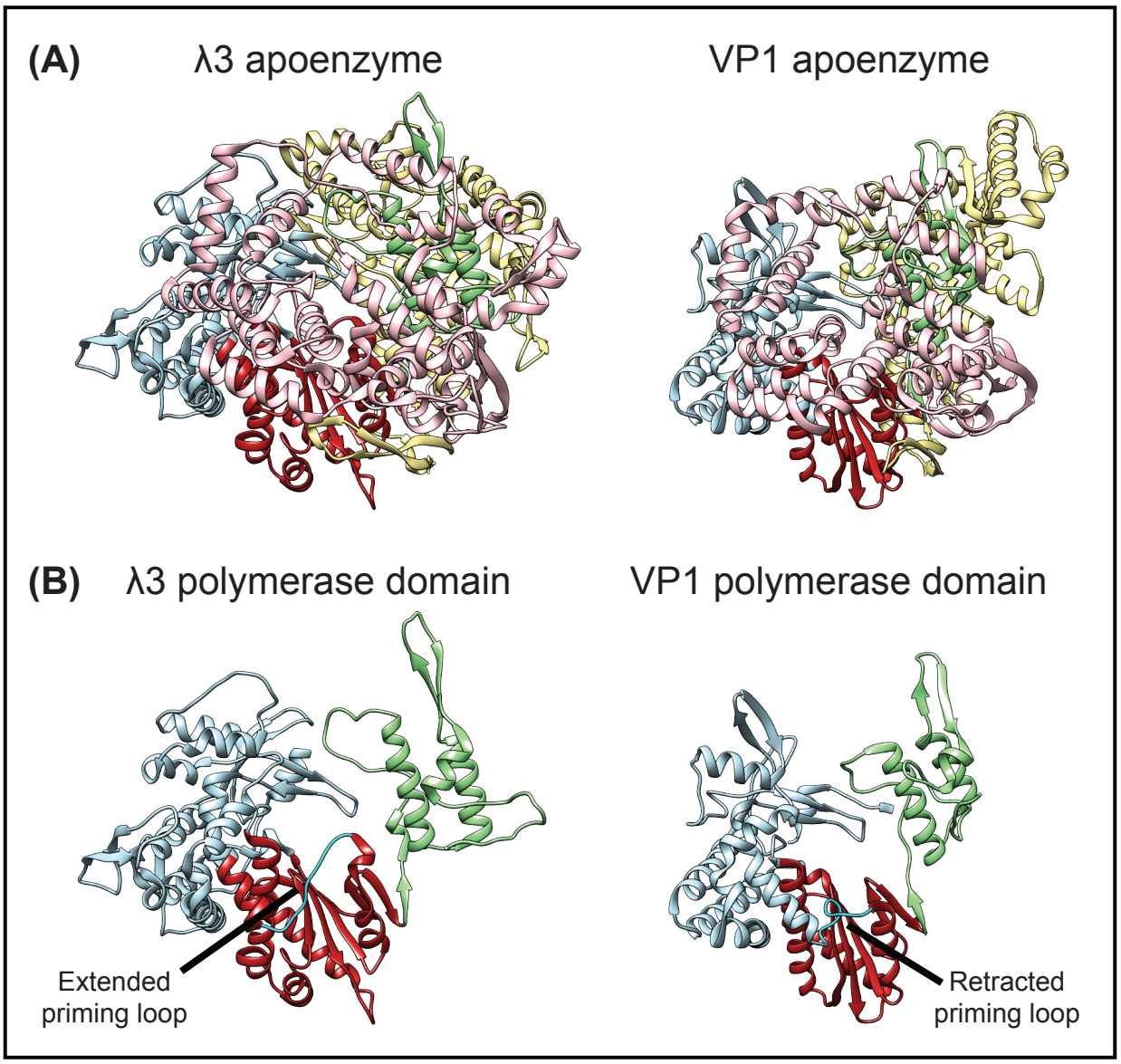


Figure 1-4. Comparison of rotavirus and reovirus RdRp structures. (A) Ribbon diagram of the reovirus λ3 (PDB accession no. 1N1H) and rotavirus VP1 (PDB accession no. 2R7Q) RdRp apoenzymes. The N- and C-terminal domains are in yellow and pink, respectively. The subdomains of the polymerase domain are in light blue (fingers), red (palm), and green (thumb). (B) The polymerase domains are shown alone for each enzyme and the priming loops are labeled and shown in cyan.

VP1 recognizes the highly conserved, single-stranded 3' terminal region of RV +RNAs (5'-UGUGACC-3'), which form an asymmetric panhandle structure that is not conserved in -RNAs, and thus only +RNAs are used as templates for dsRNA synthesis (42, 122, 124). VP1 recognition signals are distinct, however, from *cis*-acting RNA elements that promote dsRNA synthesis and there are additional recognition signals located upstream of the 3' terminal consensus sequence (42). Although VP1 can recognize and specifically bind +RNAs, these interactions alone are not sufficient to initiate dsRNA synthesis (42, 43). Moreover, addition of exogenous nucleotides, specifically GTP, is sufficient to form phosphorylated dinucleotide 'primers' (pGpG or ppGpG) that mediate dsRNA synthesis (123). Incubation with divalent cations, specifically magnesium or manganese, is essential for stabilization of newly-forming phosphodiester bonds during catalysis and also required for activation of all polymerases, including VP1 (125-127). Interestingly, initiation of dsRNA synthesis is salt-sensitive, in that replicase activity of VP1 is inhibited by low concentrations of salt, however the elongation phase of dsRNA synthesis was not affected and the mechanism underlying this salt-sensitivity is not well understood (121, 123). In all of these studies, however, VP2 remains an essential activator that is required for dsRNA synthesis, though the precise VP1-VP2 interactions have yet to be fully elucidated. Although poorly understood, VP2-dependent VP1 activation is an active and significant area of research that will be discussed in depth throughout this dissertation.

1.5G Rotavirus Virion Maturation and Egress. Nascent RV DLPs, which assemble in cytoplasmic viroplasms, traffic through the ER to acquire the outermost layer of the particle (**Fig. 1-3**). Newly-formed DLPs encounter cytoplasmic VP4 *en route* to the ER, whereby VP4 spikes associate with the VP6 shell (128). Then, interactions with NSP4 (20 kDa), a multifunctional virally-encoded ER membrane-bound receptor, facilitate budding of the particle

into the ER (129-131). NSP4-mediated budding through the ER membrane creates a unique, transient envelope surrounding the maturing RV particles. As the engulfed particles are trafficked into the interior of the ER, the envelope is quickly lost and replaced with a continuous layer of ER-resident VP7 trimers. The integrity of the outermost VP7 layer is calcium-dependent, and NSP4 additionally functions as an intracellular calcium agonist that facilitates release of calcium from intracellular stores and enables diffusion of calcium ions across the ER membrane (132, 133).

In addition to stabilizing the VP7 layer, the viroporin activities of NSP4 also trigger autophagy-related signaling cascades (134, 135). Accumulation of autophagosomes or autolysosomes at late stage infection has been associated with disruption of cellular structure, leading to lysis. However, RV virion release from infected cells is not well understood and *in vitro* studies indicate that newly-formed RV virions are capable of utilizing more than one mechanism of egress. For example, studies using nonpolarized kidney epithelial cells indicated that RVs are released via direct cell lysis (136, 137), whereas studies conducted with more biologically relevant polarized intestinal epithelial cell lines suggest that RVs are released by exocytosis secretion pathways (138). Regardless of egress mechanism, release of the RV virion from the infected cell exposes the particle to trypsin-like proteases in the gastrointestinal tract, which results in VP4 cleavage to produce the fully infectious RV virion.

Section 1.6 Host Cell Response to Rotavirus Infection

1.6A Innate Immune Response. Though the virus has evolved ways to evade or hinder immune responses by the host cell (see below), a RV-infected host cell will detect the presence of foreign viral RNA and trigger an innate immune response (139). More specifically,

differentially expressed pattern recognition receptors (PRRs) such as Toll-like receptors (TLRs), nucleotide-oligomerization domain (NOD)-like receptors, and retinoic acid-inducible gene 1 (RIG-I)-like receptors (RLRs) recognize pathogen-associated molecular patterns (PAMPs) commonly found in viruses and other intracellular microbes (140-142). PRR-PAMP interactions trigger complex signaling cascades that induce secretion of Type I interferon (IFN) cytokines and trigger antiviral interferon stimulated gene expression. Secreted IFN bind to receptors within the infected cell and on adjacent uninfected cells, and thus, interferon regulatory factors (IRFs) stimulate both autocrine and paracrine antiviral responses.

RV infection specifically triggers secretion of IFN-β and interferon stimulated gene expression of MAVS (mitochondrial antiviral signaling protein, also known as IPS-1, VISA, or Cardif) (143). MAVS signaling is induced following activation of either RIG-I and/or MDA-5 (melanoma differentiation-associated protein 5) RLRs (143-145). Although the RV-specific PAMPs recognized by RIG-I and MDA-5 are not yet known, current findings suggest that products of RV replication, rather than entering or exiting RV particles, activate RIG-I and MDA-5 (144). These cytoplasmic RLRs recruit MAVS to simultaneously activate transcription-mediated interferon regulatory factors and nuclear factor-kappa B (NF-κB) to promote interferon expression. More specifically, direct phosphorylation of cytoplasmic IRFs or phosphorylation of an NF-κB inhibitor, IκB, results in translocation of both transcription factors to the nucleus where interferon production occurs. Although IRF and NF-κB transcription factors are activated simultaneously within an RV-infected cell, they activate distinct downstream target genes: IRFs activate IFN-β pathways whereas NF-κB signaling pathways induce pro-inflammatory cytokine production (146-148).

Another cellular mechanism involved in RV-induced innate immune responses includes the 2'-5' oligoadenylate synthetase (OAS)/RNase L pathway wherein viral RNAs are degraded (149). OAS is a cytoplasmic PRR that recognizes viral dsRNA genomes and transient dsRNA replicative intermediates and subsequently synthesizes 2'-5' oligoadenylate (2-5A) products that bind and activate RNase L. RNase L subsequently cleaves viral RNAs, thereby directly inhibiting viral function. Additionally, cleaved, cytoplasmic RNAs produced by the activities of the OAS/RNase L pathway also stimulate other interferon-producing RLRs. Thus, IFN signaling induces transcription of OAS genes, which contribute to the IFN antiviral host cell environment though a positive feedback mechanism.

Though interferon production is universal in virus-infected cells, RVs have evolved immune defense mechanisms for inhibiting steps in the IFN activation pathway. For example, a virus-encoded interferon antagonist, NSP1 (59 kDa), suppresses interferon production via proteasomal degradation of IRFs (specifically IRFs 3, 5, and 7) and/or degradation of an NF-κB activator protein called β-TrCP (beta-transducin repeat containing E3 ubiquitin protein ligase) (150, 151). Moreover, VP3 antagonizes the OAS/RNase L pathway by cleaving the 2-5A OAS products, thereby precluding RNAse L activity (57, 152). Data indicates that PRR-PAMP recognition may vary based on cell type and RV strain, and there is additional, albeit incompletely understood, evidence that other PRRs such as TLR3, TLR7, and TLR9 may also be implicated in host cell response to RV infection (153, 154). In order to more completely understand the complex cell- and strain-specific roles of these factors, additional studies are warranted.

1.6B Adaptive Immune Response. Although currently licensed vaccines have shown high efficacy in preventing acute gastroenteritis, the mechanistic basis for generation of serotype

cross-reactive immunity to RV infection and disease is a poorly understood area of RV biology. Protection against RV disease is inconsistently correlated with the homotypic or heterotypic antibodies and the significance of innate immune responses for protection is unclear (155). Viral targets of protective humoral immunity are primarily directed towards VP7 and VP4 surface proteins, with some evidence for neutralizing antibodies directed towards VP6 (156-158). A near-atomic level structure of a Fab monoclonal antibody fragment binding neutralizing epitopes on VP7 has been solved, indicating that neutralizing antibodies against VP7 stabilize VP7 trimers, thus inhibiting uncoating during the RV replication cycle (159). Additional studies have revealed several other epitope binding sites on VP4 and VP7, though no high-resolution structures depicting these interactions exist yet (160-163). VP7 and VP4 elicit IgG and IgA neutralizing antibodies which react in a serotype-specific manner and correlate with protection from disease. In general, high levels of circulating serum antibody (IgG and IgA) are associated with protection, presumably via active transport mechanisms to move systemic antibody into the intestinal lumen (164, 165). Moreover, high levels of RV-specific IgA antibodies in the intestinal lumen have been associated with protection, though only a small subset are neutralizing antibodies (165, 166). *In vivo* studies in animal models have demonstrated that long-term, high-level protective immunity is dependent upon the presence of B cells, and thus antibodies, whereas cytotoxic T cells mediated short-term, low-level protection following reinfection (167, 168). Intestinal tract homing receptors for B and T cells, including $\alpha_4\beta_7$ integrin, CCR9, and CCR10, are also important for RV clearance and protection (169-171). Altogether these data suggest that correlates of RV immunity are inherently complex and may vary between species.

Section 1.7 Summary and Significance

Despite >4 decades of research, many areas of RV biology are incompletely understood. Though current vaccines have yielded significant decreases in RV-related disease burden, additional research aimed at elucidating fundamental mechanisms governing RV replication and disease are essential for the development of novel therapeutics and next-generation vaccines. Antiviral therapies targeting the RV polymerase are particularly lucrative, given that an RNA-dependent RNA polymerase (RdRp) is both virus-specific (i.e., not encoded by an uninfected host) and essential for viral replication. RV represents an ideal structurally- and functionally-controlled experimental system to gain mechanistic insight into RdRp activation and regulation due to the availability of high-resolution structures of RV particles and recombinant proteins, in addition to several well-established *in vitro* replication systems.

Synthesis of viral RNA is exquisitely regulated by a complex amalgamation of interactions that regulate VP1 activity. First, the globular VP1 structure conceals the buried catalytic core within a cage-like morphology that restricts polymerase function in the absence of specific co-factors that permeate the core through four distinct tunnels (43). For example, VP1 specifically recognizes and binds consensus sequences of RV +RNAs through an extensive network of hydrogen bonds lining the +RNA template entry tunnel (42, 43). However, even after incubation with exogenous NTPs and divalent magnesium cations, which access the buried VP1 active site via a separate exchange tunnel, the VP1/+RNA complex remains enzymatically inactive (42, 43, 118). Structural and functional studies have revealed that VP1 activity is strictly particle associated, and that the VP2 core shell is essential for VP1 activation (42, 118, 119, 172-175). Structural analyses of RV particles have revealed that VP1 is positioned off-center from the fivefold icosahedral vertices of the core shell and orientated with the +RNA exit

tunnel of VP1 abutting the core shell (55). Though it has long been confirmed that VP1 activation requires VP2, the precise mechanisms that govern VP2-dependent VP1 function are not well understood. It is not yet known if VP1 and VP2 directly bind each other to trigger VP1 activity; it is possible that VP2 binds +RNA, rather than VP1, to indirectly induce VP1 activation downstream. If VP1 and VP2 do engage each other in this context, it is not known whether these interactions are transient or stable, strong or weak. Furthermore, the VP1-VP2 interaction interface proposed by structural analyses has yet to be biochemically tested and characterized, thus it is unknown if VP1-VP2 contacts within the DLP are important for RNA synthesis. Given that VP1 functions dually as both a transcriptase and a replicase, it is possible that VP2-dependent transcriptase activity requires a different interaction interface than VP2-dependent replicase activity. Structural insights into the enzymatically inactive VP1 apoenzyme indicate that several conformational rearrangements are required for VP1 activation, though the mechanism by which these conformational changes are induced has not yet been discovered.

The work described in this dissertation addresses some of these critical gaps in knowledge and sheds light on essential regulatory mechanisms of RV polymerase function. More specifically, we investigated intergenotypic diversity of RV proteins and the effects of amino acid variability on strain-specific polymerase activity (120). Additionally, we used site directed mutagenesis and *in vitro* RNA synthesis assays to assess the functional consequences of mutating VP1-VP2 interaction interface(s) critical for RV RNA synthesis (see Chapters 3 and 4). Altogether, the results of these studies contribute valuable knowledge to the RV field and may stimulate similar analyses of RNA-dependent RNA polymerases encoded by other viruses with RNA genomes.

CHAPTER 2:

ROTAVIRUS GENOME REPLICATION: SOME ASSEMBLY REQUIRED

Courtney P. Long and Sarah M. McDonald

The published article is cited in reference #114 and may be accessed as follows:

https://doi.org/10.1371/journal.ppat.1006242.

CPL (now known as CLS) and SMM wrote the article and generated the figures.

Introduction

Viruses are obligate intracellular parasites comprised of a nucleic acid genome (RNA or DNA) that is encased within a proteinaceous capsid particle and/or lipid envelope. Genome replication and virion assembly are central processes in the lifecycles of all viruses. In many cases, a virus will first make multiple copies of its genome and then subsequently package those copies into newly formed viral capsids or envelopes. However, rotaviruses and other members of the *Reoviridae* family differ in that they replicate their genomes in concert with virion assembly. Specifically, the segmented, double-stranded (ds) RNA rotavirus genome is copied within a subviral assembly intermediate that goes on to become a mature, infectious virion. A key feature of this replicase-assembly process is that the rotavirus polymerase (VP1) is only active when tethered to the core shell protein (VP2) within the confines of an assembly intermediate. Yet several gaps in knowledge exist about the structure and composition of early

assembly intermediates for rotavirus, and the mechanism by which VP2 engages and activates VP1 is not completely understood.

Main Text

What are the structures and functions of rotavirus triple- and double-layered particles? The rotavirus virion is a triple-layered particle (TLP), ~85-nm in diameter, and it is made up of a VP2 core shell, a middle VP6 layer, and an outer VP7 layer that is embedded with VP4 spike attachment proteins (**Fig. 2-1A**) (47, 50). It is thought that the VP1 polymerase forms a complex with the VP3 RNA capping enzyme, and that VP1-VP3 heterodimers are tethered to VP2 beneath each of the 12 icosahedral fivefold axes (36). The rotavirus genome is made up of 11 dsRNA segments, coding for 6 structural proteins (VP1-VP4, VP6, and VP7) and 6 nonstructural proteins (NSP1-NSP6) (176). The TLP is the infectious form of the virus that attaches to and enters into host cells. However, during the cell entry process, the outer VP4-VP7 layer of the TLP is shed, depositing a double-layered particle (DLP) into the cell cytoplasm. VP1 polymerases within the DLP synthesize single-stranded, positive-sense (+) RNAs, which acquire a 5' cap structure (m^7GpppG) by the activities of VP3 (105). These +RNAs serve as mRNA templates for protein synthesis, and they are also selectively assorted and packaged into an early assembly intermediate where they serve as templates for genome replication by VP1 (109) (see details below). The mechanism by which rotavirus acquires one of each of its 11 genome segments remains to be determined, yet studies of other *Reoviridae* members suggest that this assortment process is mediated by RNA-RNA interactions among the single-stranded transcripts (109, 113).

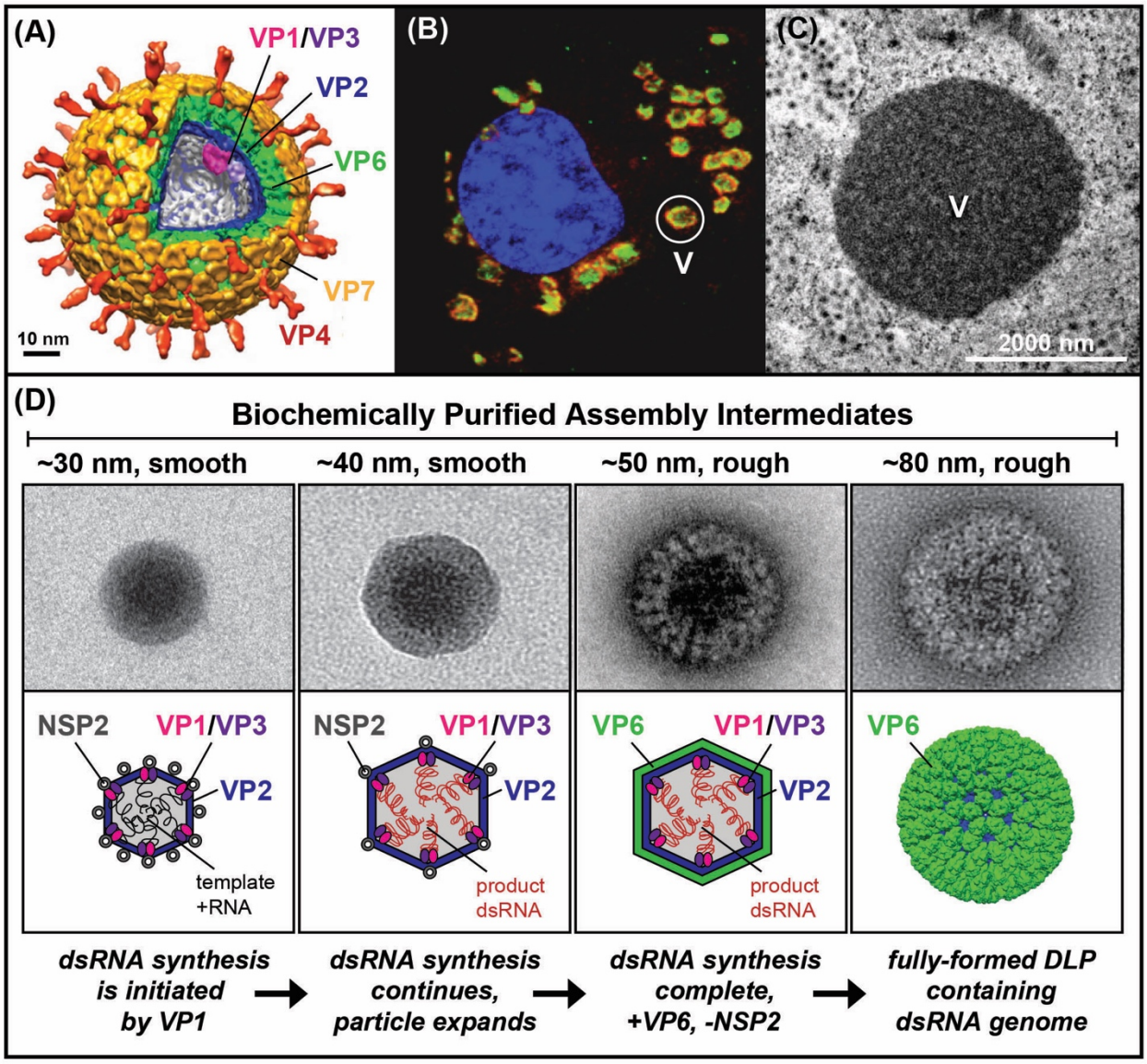


Fig 2-1. Rotavirus virion, viroplasms, and one hypothetical model of the replicase-assembly pathway. (A) Structure of the triple-layered rotavirus virion (Settembre et al 2011). Locations of viral proteins are indicated. VP1 and VP3 are modeled into the structure. (B) Confocal micrograph of a rotavirus-infected monkey kidney cell. The cell was stained to show the locations of VP2 (red), NSP2 (green), and the nucleus (blue). A single viroplasm (V) is outlined with a white circle. (C) Electron micrograph of resin-embedded, negatively-stained cell section. A single, electron-dense viroplasm (V) is shown in the center. (D) Electron micrographs of biochemically purified assembly intermediates and one hypothetical model of early rotavirus assembly. Beneath each micrograph is a cartoon representation (or, in the case of the DLP, a structure) that depicts the possible composition of the assembly intermediate, and the images are ordered according to one hypothetical pathway of early virion morphogenesis. The protein composition and activity of each particle type has not been experimentally validated.

What are the structures and functions of early rotavirus assembly intermediates?

Atomic resolution structures have been determined for rotavirus TLPs and DLPs, revealing exquisite details about capsid protein organization (47, 49, 50). In contrast, much less is known about the structures of early, replicase-competent assembly intermediates for rotavirus. One reason for our lack of knowledge about these particles is due to the fact that they are encased within viroplasms, which are discrete, cytoplasmic inclusions ~1-3 μm in diameter (**Fig. 2-1B**) (177). At <100-nm in diameter, assembly intermediates are too small to be seen using conventional light microscopy, and unfortunately, viroplasms are so electron-dense that internal features can't be resolved by higher-resolution electron microscopic (EM) imaging (**Fig. 2-1C**). Subviral particles capable of mediating *in vitro* dsRNA synthesis can be isolated from rotavirus-infected cells using biochemical approaches (87, 106, 115, 116). These putative early assembly intermediates contain VP1, VP2, VP3, and VP6 as well as NSP2, a multi-functional viral nonstructural protein critical for viroplasm formation and genome replication (178). When viewed using negative-stain EM, the isolated particles are heterogeneous in their sizes and features (**Fig. 2-1D**) (116). Specifically, the smaller particles (~30-40 nm in diameter) exhibit smooth borders, whereas the larger particles (~50-70 nm in diameter) show a rough, honeycomb pattern on their surface, which is reminiscent of 80-nm DLPs. Unlike DLPs, however, the isolated assembly intermediates are very fragile and highly permeable to metal stains and RNases, suggesting that they do not have fully-formed capsid layers. One hypothetical model of the rotavirus replicase-assembly pathway is that the smaller, smooth particles turn into the larger, rough particles and ultimately into DLPs (**Fig. 2-1D**). For instance, the ~30-nm, smooth particle could represent the earliest rotavirus assembly intermediate, within which genome replication is initiated by VP2-bound VP1. Biochemical data suggest that each VP1 polymerase functions

independently within an assembly intermediate, but that 11 polymerases act in synchrony with each other so that the genome segments are synthesized at the same time (106). It is not known how the activity of one polymerase is coordinated with those of the other polymerases. Nevertheless, as the polymerases convert the +RNAs into dsRNAs, the particle presumably expands and begins to acquire a VP6 layer, forming a larger, rough particle. Prior to final DLP assembly, the non-structural protein NSP2 must be removed. Nascent DLPs egress from the viroplasm and bud into the adjacent endoplasmic reticulum where they are converted into TLPs by addition of VP4 and VP7 (179, 180). Further studies are required to elucidate higher resolution structures, compositions, and activities of isolated rotavirus assembly intermediates and to test this hypothetical model of early morphogenesis.

How does the VP2 core shell engage the VP1 polymerase during early particle assembly? To initiate genome replication (i.e., dsRNA synthesis) the VP1 polymerase must be bound by the core shell protein VP2. In DLPs or TLPs, VP2 is organized as 12 interconnected, decameric units (**Fig 2-2A**), but its structural organization in assembly intermediates is not known (47, 49, 50). The VP2 monomers in each decamer unit adopt one of two slightly different conformations (VP2-A and VP2-B). One conformation, VP2-A, converges tightly around the icosahedral axis, whereas the other conformation, VP2-B, intercalates between adjacent VP2-A monomers. The extreme N-terminal region of VP2 (residues ~1-100) protrudes inward and makes contact with VP1 (**Fig. 2-2B**). Estrozi et al. predicted that VP1 is positioned against the inner surface of the VP2 core shell off-center from the five-fold axis and that it is stabilized against the core shell by VP2 N termini (55). The regions of VP2 that contact VP1 in the DLP structure are the same as those shown to be important for VP2-mediated VP1 enzymatic

activation *in vitro* (119). This observation suggests that the VP1 and VP2 binding interaction during the early replicase-assembly process may be similar to the VP1 and VP2 binding interaction during transcription.

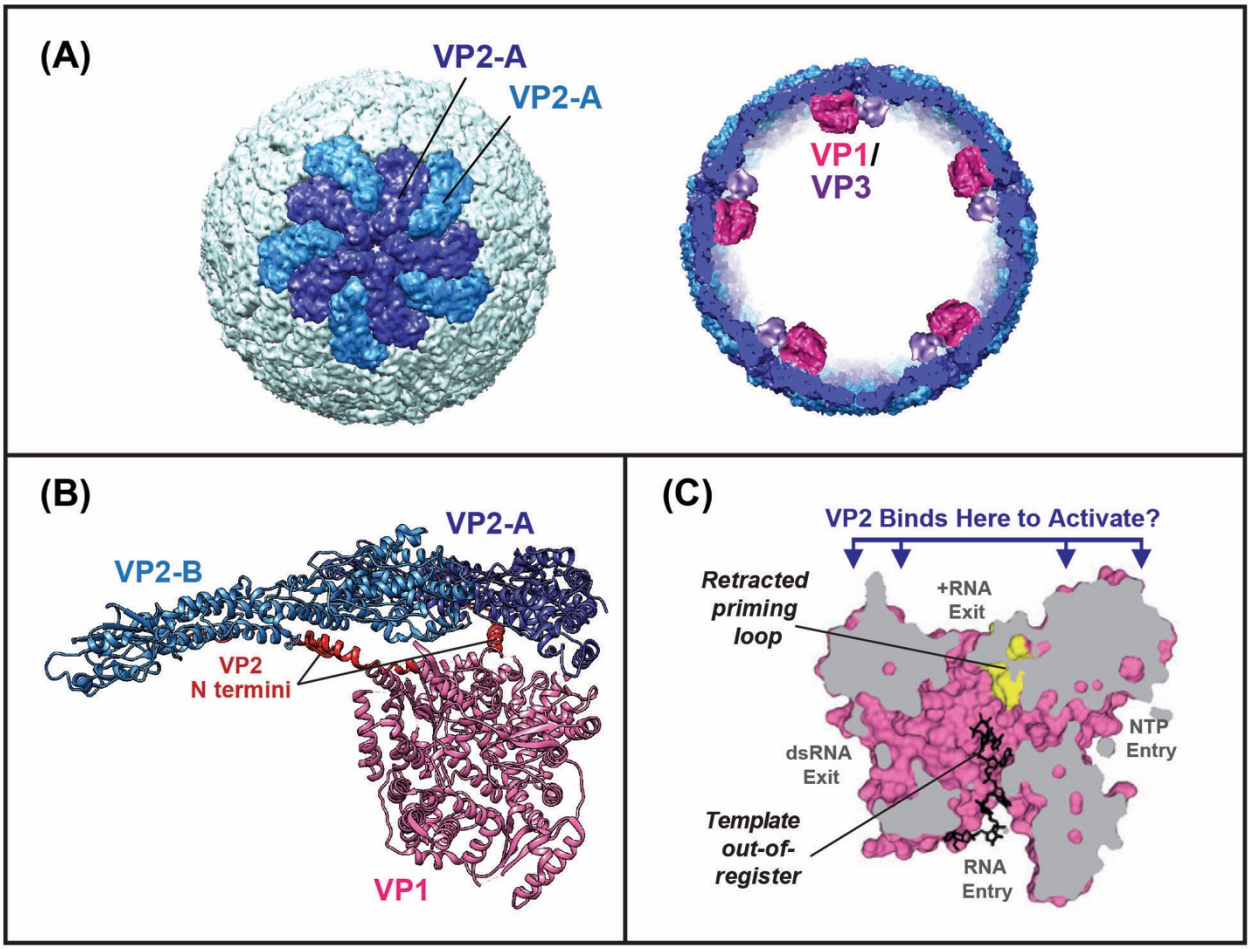


Fig 2-2. Rotavirus VP1-VP2 interactions. (A) Left: Structure of the rotavirus VP2 core shell in the context of a DLP (Lawton et al 2000). VP2-A and VP2-B monomers of one decamer unit are colored blue and cyan, respectively. Right: The core shell is shown computationally sliced through the middle to reveal the likely locations of VP1-VP3 heterodimers. (B) VP1-VP2 contacts in the fully-assembled DLP (Estrozi et al 2013). VP2-A, VP2-B, and VP1 are shown in ribbon representation and colored blue, cyan, and pink, respectively. The visible regions of the VP2 N termini are colored red. (C) VP1 is shown in surface representation and computationally sliced through the middle to reveal four tunnels (Lu et al 2008). The location of the retracted priming loop is shown in yellow, and a 7-nucleotide +RNA template that is bound out-of-register with the active site is shown in black. VP2 is predicted to engage VP1 at the +RNA exit interface and trigger structural rearrangements including a repositioning of the priming loop and +RNA template.

VP1 is a globular, cage-like enzyme comprised of a central polymerase domain (with canonical finger, palm, and thumb subdomains) that is flanked by extended N- and C-terminal domains (43). Together, these three domains create four distinct tunnels within VP1 that support NTP entry, +RNA entry, +RNA exit, and dsRNA exit (**Fig. 2-2C**). Within the DLP, VP1 is proposed to be oriented with its +RNA exit tunnel facing towards the VP2 core shell; such an orientation would facilitate the egress of +RNAs from the DLP capsid layers during transcription (**Fig. 2-2C**) (105). This orientation of VP1 is also similar to that proposed for the polymerases of other *Reoviridae* family members (38, 117). However, the importance of VP2 binding to the VP1 +RNA interface for VP1 replicase function has yet to be biochemically tested. It also remains unclear how VP2 engagement of VP1 leads to enzymatic activation of the polymerase. The structure of VP1 in complex with template +RNA—but in the absence of VP2—reveals an auto-inhibited, inactive polymerase with at least two mal-positioned elements: (i) the so-called 'priming loop' of VP1, which stabilizes the initiating NTP during genome replication, is bent too far forward, and (ii) the +RNA template is bound out-of-register with the active site (**Fig. 2-2C**) (43). Thus, VP2 engagement of VP1 may trigger a series of conformational changes in the polymerase interior including repositioning the priming loop and +RNA template to allow for the initiation of dsRNA synthesis. Further studies seeking to test this hypothesis will benefit from a robust *in vitro* assay that recapitulates VP2-dependent VP1 activation (42, 181).

Perspectives

During their intracellular lifecycles, viruses make numerous copies of their nucleic acid genomes and package them into nascent particles. Viral genome replication and particle assembly are often highly coordinated within the infected cell to maximize efficiency.

Rotaviruses and other *Reoviridae* family members may very well exhibit the utmost level of coordination, as they replicate their genomes concurrent with assembly of new virions. The mechanism of this concerted replicase-assembly process is incompletely understood. Isolated rotavirus subviral particles that can perform dsRNA synthesis *in vitro* are just beginning to be characterized in terms of their structure and composition, and there is much to be learned about how the activity of the rotavirus VP1 polymerase is regulated via interaction(s) with the core shell protein VP2 in the context of the assembling particle. Although other viruses do not perform this same multi-tasking feature of replicating their genomes while assembling particles, it is apparent that they also must regulate the activities of their polymerases. The vast majority of viral polymerases do not function as sole polypeptides during infection (182). Instead, they are components of multi-subunit complexes, and interactions among the protein constituents dictate the polymerase activity. Thus, studies of rotavirus polymerase regulation during particle assembly may broadly inform an understanding of how other viruses ensure that genome replication occurs at the right place and time in the infected cell.

CHAPTER 3:

GROUP A ROTAVIRUS VP1 POLYMERASE AND VP2 CORE SHELL PROTEINS: INTERGENOTYPIC SEQUENCE VARIATION AND *IN VITRO* FUNCTIONAL COMPATIBILITY

Courtney L. Steger, Crystal E. Boudreaux, Leslie E. LaConte, James B. Pease, and Sarah M. McDonald

The published manuscript is cited in reference #120 and may be accessed as follows:

https://jvi.asm.org/content/jvi/93/2/e01642-18.full.pdf

CLS and SMM wrote the manuscript and generated the figures. CEB performed preliminary replicase assays that contributed to the development of this project. LEL performed the molecular dynamics simulations for VP1. JBP performed multidimensional scaling analyses for VP1 and VP2.

Abstract

Group A rotaviruses (RVAs) are classified according to a nucleotide sequence-based system that assigns a genotype to each of the 11 double-stranded (ds) RNA genome segments. For the segment encoding the VP1 polymerase, 22 genotypes (R1-R22) are defined with an 83% nucleotide identity cut-off value. For the segment encoding the VP2 core shell protein, which is a functional VP1-binding partner, 20 genotypes (C1-C20) are defined with an 84% nucleotide identity cut-off value. However, the extent to which the VP1 and VP2 proteins encoded by these

genotypes differ in their sequences or interactions has not been described. Here, we sought to (i) delineate the relationships and sites of variation for VP1 and VP2 proteins belonging to the known RVA genotypes and (ii) correlate intergenotypic sequence diversity with functional VP1-VP2 interaction(s) during dsRNA synthesis. Using bioinformatic approaches, we revealed which VP1 and VP2 genotypes encode divergent proteins and identified the positional locations of amino acid changes in the context of known structural domains/subdomains. We then employed an *in vitro* dsRNA synthesis assay to test whether genotype R1, R2, R4, and R7 VP1 polymerases could be enzymatically-activated by genotype C1, C2, C4, C5, and C7 VP2 core shell proteins. Genotype combinations that were incompatible informed the rational design and *in vitro* testing of chimeric mutant VP1 and VP2 proteins. The results of this study connect VP1 and VP2 nucleotide-level diversity to protein-level diversity for the first time, and they provide new insights into regions/residues critical for VP1-VP2 interaction(s) during viral genome replication.

Importance

Group A rotaviruses (RVAs) are widespread in nature, infecting numerous mammalian and avian hosts and causing severe gastroenteritis in human children. RVAs are classified using a system that assigns a genotype to each viral gene according to its nucleotide sequence. To date, 22 genotypes have been described for the gene encoding the viral polymerase (VP1) and 20 genotypes for the gene encoding the core shell protein (VP2). Here, we analyzed if/how the VP1 and VP2 proteins encoded by the known RVA genotypes differ from each other in their sequences. We also used a biochemical approach to test whether the intergenotypic sequence differences influenced how VP1 and VP2 functionally engage each other to mediate RNA

synthesis in a test tube. This work is important because it increases our understanding of RVA protein-level diversity and raises new ideas about VP1-VP2 binding interface(s) that are important for viral replication.

Introduction

Rotaviruses are members of the *Reoviridae* family that package an 11-segmented double-stranded (ds) RNA genome within a non-enveloped, icosahedral particle (176). More than 8 species of rotavirus have been identified in nature (groups A-H, and tentative groups I and J), though the vast majority of human and animal disease is caused by group A rotavirus (RVA) strains (30-32). RVAs are economically and medically significant gastrointestinal pathogens that remain important causes of childhood diarrhea, leading to ~215,000 deaths each year (6). RVAs also infect a wide range of mammalian and avian hosts including, but not limited to, cows, pigs, dogs, cats, chickens, turkeys, pigeon, mice, bats, horses, monkeys, and rabbits (176). Individual RVA strains are classified according to a system that assigns a specific genotype to each of the 11 genome segments according to its sequence and established nucleotide percent identity cut-off values (34) (**Table 3-1**). In most strains, each genome segment is a gene that encodes a single viral protein. Thus, 6 genes encode the structural proteins VP1-VP4, VP6, and VP7 and 5 genes encode the non-structural proteins NSP1-NSP5(176). In the classification system, each gene is designated by a single letter (defined by the encoded protein function), and the assigned genotype is listed as a number (**Table 3-1**). To date, 20-51 different genotypes have been described for RVA genes (35), indicating a high degree of nucleotide-level strain diversity (**Table 3-1**). However, the extent to which this diversity is reflected in the sequences, structures, or functions of the encoded proteins has not yet been described.

The rotavirus RNA-dependent RNA polymerase, VP1, is the largest protein encoded by RVAs (125-kDa; ~1090 amino acids in length). Currently, 22 different VP1 genotypes (R1-R22) have been defined for RVAs based upon an 83% nucleotide identity cut-off value for the gene (34) (**Table 3-1**). While previous studies have investigated the amino acid sequence conservation and variation for VP1 proteins belonging to different rotavirus groups (i.e., RVA-RVD and RVF-RVH), much less is known about if or how the RVA VP1 proteins encoded by genotypes R1-R22 differ from each other (44, 183). The structure of a genotype R2 VP1 protein (strain SA11) has been solved to atomic resolution, revealing a globular, cage-like enzyme with an N-terminal domain (NTD; amino acids ~1-332), a central polymerase domain (amino acids ~333-778), and a C-terminal 'bracelet' domain (CTD; amino acids ~779-1089) (**Fig. 3-1A**) (43). The catalytic center resides within the polymerase domain and is comprised of several

Gene/Protein	% NT Identity	Genotypes	Function(s)
VP1	83	R1-R22	**R**NA-dependent RNA polymerase
VP2	84	C1-C20	**C**ore shell protein
VP3	81	M1-M20	**M**ethyltransferase
VP4	80	P[1]-P[51]	**P**rotease-sensitive attachment
VP6	85	I1-I26	**I**ntermediate capsid layer
VP7	80	G1-G36	**G**lycoprotein
NSP1	79	A1-A31	Innate immune **A**ntagonist
NSP2	85	N1-N22	**N**TPase; viroplasm formation
NSP3	85	T1-T22	**T**ranslation enhancer
NSP4	85	E1-E27	**E**ntertoxin; assembly
NSP5/6	91	H1-H22	P**H**osphoprotein; viroplasm formation

Table 3-1: Genotype Classification of Individual RVA Genome Segments

structurally-conserved motifs, all of which are buried in the hollow interior of the enzyme. Four tunnels permeate VP1 and allow for the entry of single-stranded RNA templates, NTPs, and divalent cations, as well as the exit of RNA templates and positive-sense (+) RNA transcripts or dsRNA products of genome replication (**Fig. 3-1B**). Despite having all the necessary structural features for catalysis, VP1 is not active as an apoenzyme. Instead, VP1 function is dependent upon its interaction(s) with VP2, the core shell protein that comprises the innermost layer of the triple-layered virion (179, 184).

VP2 is the second largest protein encoded by RVAs (102-kDa; ~900 amino acids in length), and 20 different genotypes (C1-C20) have been defined based upon an 84% nucleotide sequence identity cut-off value for the gene (34) (**Table 3-1**). Yet, the extent to which VP2 proteins encoded by the 20 different genotypes differ has not yet been described. Atomic structures of VP2 in the context of triple-layered virions and double-layered particles (DLPs) have been solved for strain SA11 (genotype C5) and strain UK (genotype C2) (47, 49, 50). All structures reveal a smooth, T=1 icosahedral shell comprised of 120 VP2 monomers, arranged into 12 decameric units (**Fig. 3-1C**). VP2 monomers adopt two slightly different conformations, with one conformation (VP2-A) tightly enclosing the fivefold axis and the other (VP2-B) sitting further away from the fivefold and interdigitating VP2-A. The principal scaffold domain of VP2 (amino acids ~101-900) is made up of apical, carapace, and dimer-forming subdomains (**Fig. 3-1D**). The extreme N-terminal domain (NTD) of VP2 (amino acids ~1-100) is not fully resolved in any of the structures. However, all available data indicates that the VP2 NTD protrude into the core interior (49, 50, 119).

The RVA polymerase, VP1, performs both transcription (+RNA synthesis) and genome replication (dsRNA synthesis) in connection with the VP2 core shell (43, 114, 183). Yet, the

manner in which these proteins engage each other during each stage of RNA synthesis is not fully understood. Transcription occurs in the context of a DLP, wherein VP1 monomers are oriented beneath the fivefold icosahedral vertices such that their +RNA exit tunnels abut the VP2 capsid layer (55, 105) (**Fig. 3-1E**). Estrozi et al. showed that regions of the VP1 NTD and CTD are contacted by both the VP2 NTD and principal scaffold domain in the DLP structure of the bovine strain UK (VP1 genotype R2 and VP2 genotype C2) (55). Less is known about the position and orientation of VP1 during genome replication, which occurs within subviral, assembly-replicase intermediates (87, 106, 115, 116). There are currently no high-resolution structures of VP1 and VP2 within these subviral particles. The little that we do know about the VP1-VP2 interaction(s) during genome replication is based upon the results of *in vitro* dsRNA synthesis assays mostly using recombinant proteins of strain SA11 (genotype R2 VP1 and genotype C5 VP2) (119). These results suggest that an intact VP2 NTD is essential for supporting robust rVP1 activity and also that regions of the VP2 principal scaffold domain mediate VP1 interaction(s) in a group-specific manner (119, 183). More specifically, it was shown that residues in the VP2 apical and/or carapace subdomain, but not the NTD, allow the core shell protein to differentiate between a VP1 from RVA (strain SA11) vs. RVC (strain Bristol) (119). *In vitro* assays have also been used to probe regions/residues of VP1 that are important for dsRNA synthesis, particularly those involved in template binding and catalysis (44, 185, 186). Still, the regions of VP1 that functionally interact with VP2 during *in vitro* dsRNA synthesis are not known.

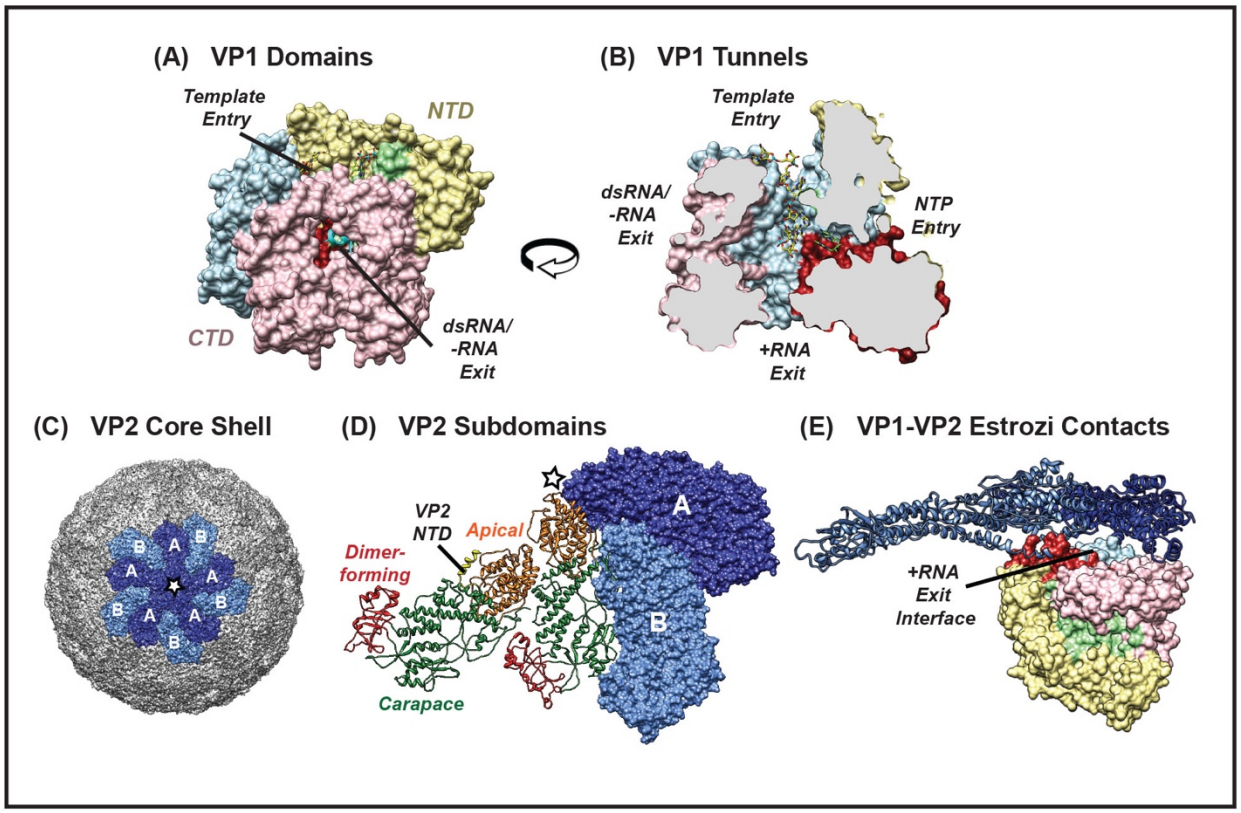


Fig. 3-1. Structures of VP1 and VP2. (A) Atomic structure of SA11 (R2) rVP1 (PDB accession no. 2R7R) is shown in surface representation and colored according to domain/subdomain organization. The N-terminal domain (NTD; yellow) and C-terminal domain (CTD; pink) surround the central polymerase domain, which is comprised of canonical finger (blue), palm (red), and thumb (green) subdomains. The extreme C terminus of the protein forms a plug, which is colored in cyan and shown in ribbon representation. (B) Structure in panel A is shown rotated 90° to the left and computationally sliced through the middle to reveal four tunnels extending into the catalytic core. Known or putative functions of the tunnels are labeled, and a 7-nucleotide RNA template is shown in stick figure representation. (C) Structure of the core shell of a bovine rotavirus (PDB accession no. 3KZ4) depicted in surface representation. A central VP2 decamer is colored in dark and light blue to highlight five VP2-A and five VP2-B monomers, respectively. (D) Two neighboring VP2 dimers are shown (each comprised of one VP2-A monomer and one VP2-B monomer). One dimer is depicted in surface representation, with the same coloration as in panel C. The other VP2 dimer is shown in ribbon representation and colorized to show subdomain organization of the principal scaffold domain (red: dimer-forming; orange: apical; and green: carapace). The resolved portion of the VP2 NTD is colored in yellow and indicated on the structure. (E) Structure depicting VP1-VP2 contacts within the bovine RVA DLP (PDB accession no. 4F5X). VP1 and VP2 are colored as in panels A-C. The VP1 monomer is oriented beneath the VP2 capsid layer such that residues surrounding the +RNA exit tunnel contact VP2.

In this study, we sought to investigate the sequences and *in vitro* functions of genetically-diverse RVA VP1 and VP2 proteins. Specifically, we used bioinformatic approaches to determine how similar/different are VP1 and VP2 proteins encoded by the known RVA genotypes, thereby connecting nucleotide-level diversity to protein-level diversity for the first time. Furthermore, we mapped regions of conservation and variation within the polymerase and core shell proteins as well as identified amino acid positions that distinguish certain VP1 and VP2 genotypes/lineages. Moreover, using *in vitro* dsRNA synthesis assays, we tested the functional compatibility of wildtype and mutant VP1 and VP2 proteins belonging to several different genotypes. Our results provide a comprehensive description of RVA VP1 and VP2 intergenotypic amino acid sequence diversity and shed light on how this diversity might correlate with functional VP1-VP2 interactions during genome replication.

Results

Intergenotypic relationships among RVA VP1 genes and proteins. Currently, 22 different VP1 genotypes (R1-R22) have been defined for RVAs based upon an 83% nucleotide identity cut-off value (34) (**Table 3-1**). We first sought to determine the relationships among these VP1 genotypes at the nucleotide level. To do this, VP1 open-reading frame (ORF) nucleotide sequences were retrieved from public databases, and representatives of each genotype were used to construct a maximum-likelihood phylogenetic tree (**Appendix A and Fig. 3-2A**). Sequences for genotypes R17, R21, and R22 were not available at the time of the study. The results of this analysis revealed two major lineages (I and II) of VP1 genotypes that were differentiated by mammalian vs. avian hosts. More specifically, lineage I was comprised of 16 VP1 genotypes (R1-R3, R5, R7-R13, R15, R16, and R18-R20), which were exclusively found in

strains isolated from mammals (**Appendix A and Fig. 3-2A**). Lineage II, on the other hand, was comprised of 3 VP1 genotypes (R4, R6, and R14), which were mostly found in strains isolated from avian hosts (turkey, pheasant, chicken, duck, pigeon, and gull). One exception for the avian-specificity of lineage II was a single R4 genotype from a strain that was isolated from a fox, possibly following an inter-species transmission event (187) (**Appendix A**).

For mammalian RVA lineage I, genotypes R7, R13, R15, R16, R19, and R20 were cleanly delineated in the phylogenetic tree with strong bootstrap support values (**Fig. 3-2A**). Genotype R7 has been found exclusively in murine RVA strains that are commonly used in laboratory studies (188) (**Appendix A**). Genotypes R15, R16, R19, and R20 were all found in modern RVAs isolated from bats (189, 190) (**Appendix A**). Genotype R13 also likely originated from a bat RVA, but the only representative was isolated from a 2-yr old child in Suriname, presumably via an interspecies transmission event (191) (**Appendix A**). The remaining lineage I genotypes (R1-R3, R5, R8-R12, and R18) formed distinct phylogenetic groupings but lacked strong bootstrap support values (**Fig. 3-2A**). Genotype R1 was found in strains isolated from humans and pigs, with only a few exceptions (**Appendix A**). Genotype R2 was found in strains isolated from humans and numerous other animal hosts including, but not limited to, cows, sheep, cats, giraffes, pigs, deer, rabbits, horses, camels, goats, antelopes, monkeys, and bats (**Appendix A**). These two genotypes (R1 and R2) are typically associated with human strains that cause pediatric gastroenteritis (176, 192, 193). Genotype R3 was also found in strains isolated from humans; however, this genotype is only rarely associated with gastroenteritis in children. Genotype R3 is most likely of animal RVA origin, and it was found in strains isolated from several different hosts including, but not limited to dogs, rabbits, horses, bats, and monkeys (**Appendix A**). Genotypes R5 and R9 are additional animal RVA genotypes that show

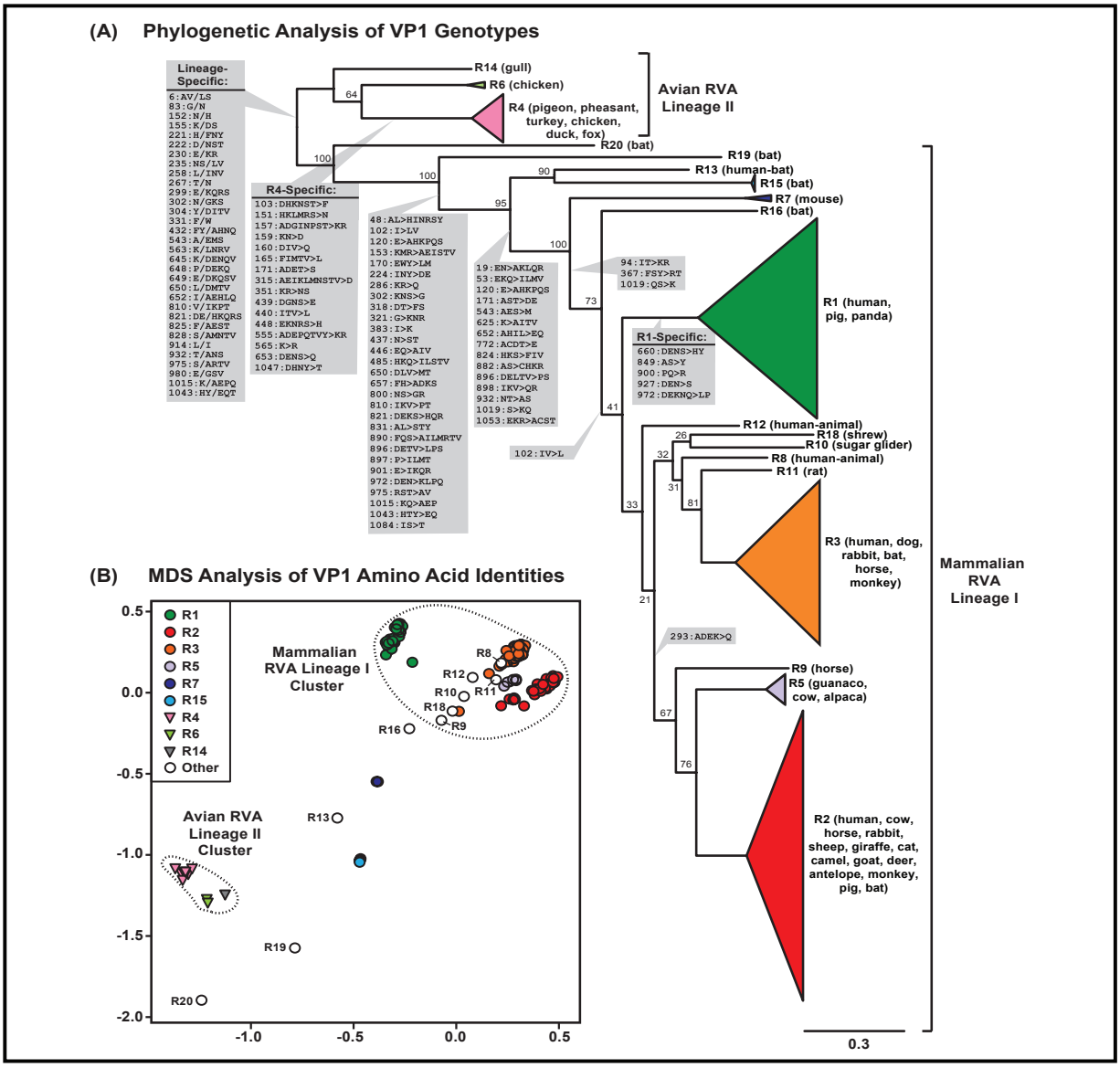


Fig. 3-2. Intergenotypic relationships among VP1 genes and proteins. (A) An intergenotypic maximum-likelihood phylogenetic tree was inferred from the aligned ORFs of 158 sequences representing genotypes R1-R16 and R18-R20. Horizontal branch lengths are drawn to scale (nucleotide substitutions per base) with bootstrap values shown as percentages for key nodes. Monophyletic groupings were collapsed and are shown as colored, cartooned triangles representing a single genotype. Genotypes are indicated for each branch/triangle, and the animal host of the corresponding RVA is listed in parentheses. Brackets indicate two phylogenetically-distinct lineages of VP1 proteins (I and II). Amino acid positions and residue changes that differentiate key branch points are overlaid on the tree (gray boxes). In all cases, the amino acid change(s) is listed in the same directionality as the tree (i.e., amino acid(s) of left branches>amino acids of right branches. Position number is based upon that of the longest gene. (B) Pairwise amino acid sequence distances are represented in an MDS plot. X- and Y- axes are arbitrary coordinates in 2D space that allowed proper visualization of genotypes. Colors of genotypes are the same as in panel A, and clusters are indicated by dotted lines.

phylogenetic relationships with each other and with the R2 genotype. Genotypes R8, R10-12, and R18 were found in various animal hosts, and human strains with genotypes R8 and R12 are suspected inter-species transmission events (194, 195) (**Appendix A**). Genotypes R8, R10-12, and R18 seem to share a phylogenetic relationship with each other and with the R3 genotype.

Having observed the relationships of VP1 genotypes at the nucleotide level, we next sought to determine if and/or how these relationships are reflected in the sequences of the encoded proteins. To do this, we constructed an intergenotypic amino acid sequence alignment and calculated all pairwise sequence identity values. This data was then used in a multi-dimensional (MDS) scaling analysis, which is a multi-variate procedure designed to visualize and identify patterns in a distance matrix (196) (**Fig. 3-2B**). In MDS, sequences are represented as points whose respective pairwise sequences distances are approximated when projected onto a two-dimensional (2D) plane. Thus, clusters of closely-spaced points represent sequences that cannot be readily distinguished from each other based upon their pairwise distances. In contrast, points that are spread far apart in planar space represent proteins that are more divergent in their pairwise sequence distances. The MDS results for VP1 show that the points representing avian RVA lineage II proteins clustered distinctly from those of mammalian RVA lineage I proteins (**Fig. 3-2B**). Within the avian RVA lineage II cluster, there was also strong intragenotypic point overlap (e.g., all the R4 pink triangles formed a tight overlapping cluster) and intergenotypic point differentiation (e.g., the R4 pink triangles did not overlap with the R6 light green triangles) (**Fig. 3-2B**). This result suggests that the VP1 proteins encoded by avian RVA lineage II genotypes R4, R6, and R14 are sufficiently different from each other and from those of mammalian RVA lineage I (R1-R3, R5, R7-R16, and R18-R20). In contrast, points representing the mammalian RVA lineage I VP1 proteins did not mutually cluster, and there was strong

overlap among several of the genotypes. More specifically, a cluster containing the common genotypes R1, R2, and R3 also included several rare genotypes (R5, R8-12, and R18) (**Fig. 3-2B**). Points representing genotypes R1 and R2 (green and red circles, respectively) remained cleanly separated from each other and from those of other genotypes in this cluster, suggesting strong genotype-specific definition for these VP1 proteins in regard to their sequence identities. In contrast, points representing genotype R3 (orange circles) showed more diffuse clustering with points that spatially-overlapped with those for R5, R8-12, and R18 genotypes. This result suggests that R3, R5, R8-12, and R18 VP1 proteins are not easily distinguishable from each other based upon pairwise amino acid sequence distances. On the other hand, points representing mammalian RVA lineage I genotypes R7, R13, R15, R16, R19, and R20 were spatially separate from each other and from those the other avian/mammalian RVA genotypes (**Fig. 3-2B**). Thus, these VP1 proteins are highly divergent in their amino acid sequences.

Given the observation that some, but not all, of the genotypes could be differentiated by pairwise distance in the MDS analysis, we next sought to determine whether any of the VP1 genotypes exhibit specific amino acid signatures that can be used to identify them (**Fig. 3-2A**). In particular, we wanted to identify amino acid positions that change in concordance with phylogenetic groupings, which in this case includes genotypes (197). The results for VP1 show that only R1 and R4 had genotype-specific amino acid signatures in their VP1 proteins (**Fig. 3-2A**). All other amino acid positional changes were shared among several different genotypes. In particular, 32 amino acid positions were identified that differentiated mammalian RVA lineage I vs. avian RVA linage II VP1 proteins. Also, 5 branch points that fall within lineage I correlated with amino acid signatures, suggesting that these residues are shared among several different mammalian RVA VP1 genotypes (**Fig. 3-2A**).

Intergenotypic relationships among RVA VP2 genes and proteins. We next used the same approaches described above to determine the relationships among the VP2 genotypes C1-C20. ORF nucleotide sequences for genotypes C16, C19, and C20 were not available in the databases at the time of this study. Similar to what was found for VP1, two major lineages (I and II) were seen in the intergenotypic maximum-likelihood phylogenetic tree for VP2 (**Fig. 3-3A**). Lineage II was comprised of genotypes C4, C6, and C14, which were mostly found in strains isolated from avian hosts (turkey, pheasant, chicken, duck, pigeon, and gull), with the exception of a C4 VP2 genotype from the same fox strain that had an R4 VP1 genotype (187) (**Appendix A**). Lineage I was comprised of the remaining 14 VP2 genotypes (C1-C3, C5, C7-C13, C15, C17, and C18), which were exclusively found in strains isolated from mammalian hosts (**Appendix A and Fig. 3-3A**). C13 was the only mammalian RVA lineage I genotype that could be cleanly delineated in the phylogenetic tree with strong bootstrap support (**Fig. 3-3A**). This genotype was found from the same bat RVA-like human strain that contained R13 (191). The remaining VP2 genotypes formed separate groupings in the tree, but they did not show strong bootstrap support values (**Fig. 3-3A**). Genotype C1 VP2 was found mostly along with R1 VP1 in strains isolated from humans and pigs, with only a few exceptions (**Appendix A**). Genotype C2 VP2 was associated with R2 or R3 VP1 in strains isolated from humans and numerous other animal hosts (**Appendix A**). Genotype C3 VP2 was associated with R2 or R3 VP1 and also found in strains of likely animal RVA origin (**Appendix A**). Genotype C7 was found associated R7 VP1 in murine RVA strains (188) (**Appendix A**). Genotypes C8, C15, and C18 were found in modern RVAs from bats along with R16, R15, and R19 VP1 genotypes, respectively (189, 190) (**Appendix A**). VP2 genotypes C9, C10, C11, and C12 were from various animals and associated with R9, R10, R11, R12, VP1 genotypes, respectively (**Appendix A**).

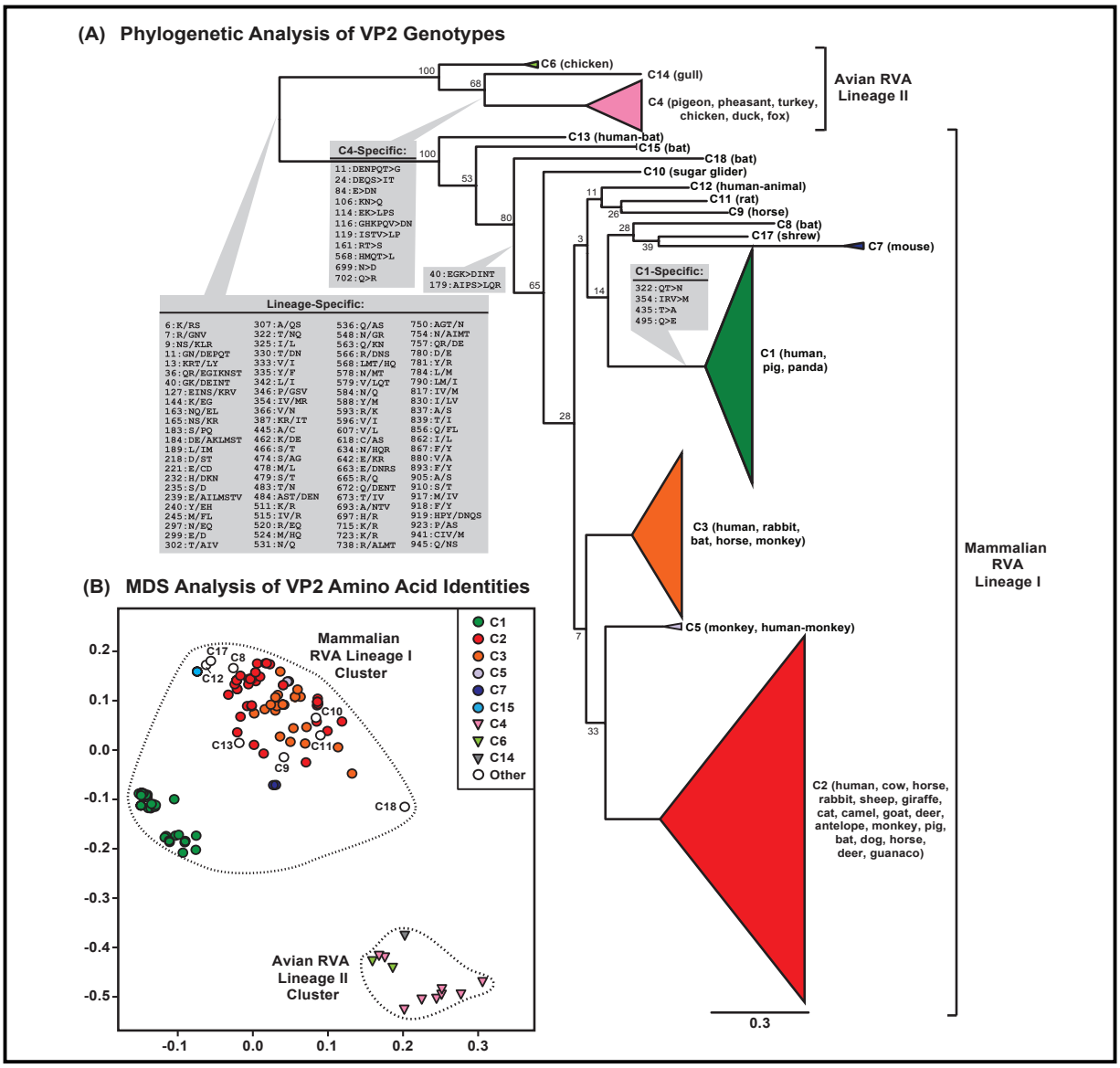


Fig. 3-3. Intergenotypic relationships among VP2 genes and proteins. (A) An intergenotypic maximum-likelihood phylogenetic tree was inferred from the ORF data for 158 sequences representing genotypes C1-C15 and C17-C18. Horizontal branch lengths are drawn to scale (nucleotide substitutions per base) with bootstrap values shown as percentages for key nodes. Monophyletic groupings were collapsed and are shown as colored, cartooned triangles representing a single genotype. Genotypes are indicated for each branch/triangle, and the animal host of the corresponding RVA is listed in parentheses. Brackets indicate two phylogenetically-distinct lineages of VP2 proteins (I and II). Amino acid positions and residue changes that differentiate key branch points are overlaid on the tree (gray boxes). In all cases, the amino acid change(s) is listed in the same directionality as the tree (i.e., amino acid(s) of left branches>amino acids(s) of right branches. Position number is based upon that of the longest gene. (B) Pairwise amino acid sequence distances between are represented in an MDS plot. X- and Y- axes are arbitrary coordinates in 2D space that allowed proper visualization of genotypes. Colors of genotypes are the same as in panel A, and lineages are indicated by dotted lines.

The MDS analysis for VP2 revealed that the points representing avian RVA lineage II genotypes (C4, C6, and C14) clustered distinctly from those representing the mammalian RVA lineage I VP2 genotypes (C1-C3, C5, C7-C13, C15, C17, and C18) (**Fig. 3-3B**). However, unlike the tight differentiation that was seen with avian RVA VP1 genotypes, the points representing avian RVA VP2 genotypes C4 (pink triangles) and C6 (light green triangles) mutually colocalized with each other, suggesting that they are indistinguishable based upon their pairwise sequence distances. Likewise, there was relatively close clustering among the mammalian RVA lineage I genotypes with the exception of C1, C7, and C18, which showed strong separation (**Fig. 3-3B**). In particular, the points representing C2, C3, C8-C13, C15, and C17 showed spacing that was not congruent with genotype designation, suggesting that these proteins are not easily differentiated from each other. The branch-specific amino acid analysis for VP2 revealed genotype-specific amino acid signatures for C1 and C4 (**Fig. 3-3A**). Moreover, 96 amino acid signatures were found to differentiate lineage I vs. II VP2 proteins, and 2 positional changes appeared to be shared among several mammalian RVA lineage I genotypes (i.e., C18 branch split) (**Fig. 3-3A**).

Conserved and variable regions of RVA VP1 and VP2 proteins. Having investigated the intergenotypic relationships among RVA VP1 and VP2 proteins from a bioinformatics perspective, we next sought to analyze the domains, subdomains, and motifs, thereby identifying regions/residues of conservation and variation among genotypes. To do this, we generated amino acid sequence alignments using consensus sequences for each available genotype of VP1 or VP2 (**Figs. 3-4, 3-5 and Appendices B-C**). The VP1 alignment revealed that sequence variation was distributed across the entire protein (**Fig. 3-4 and Appendix B**). Although R20 clustered phylogenetically with mammalian RVA lineage I genotypes, the amino acid sequence

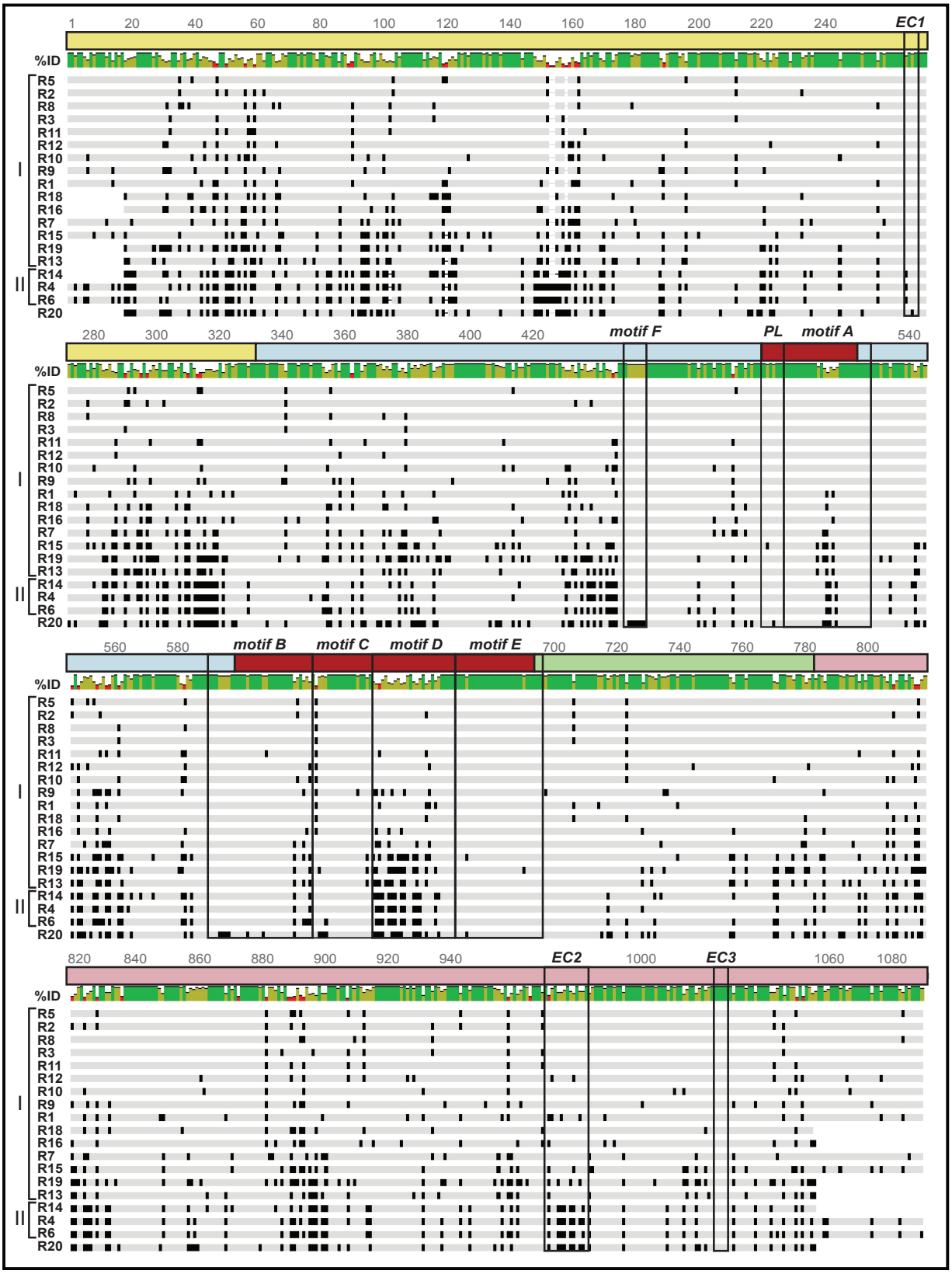


Fig. 3-4. VP1 genotype consensus amino acid sequence alignment. Legend on following page.

Fig. 3-4. VP1 genotype consensus amino acid sequence alignment. The schematic shows an amino acid sequence alignment of RVA VP1 consensus sequences for genotypes R1-R16 and R18-R20. The VP1 domains and subdomains are represented by a line above the sequence and are colored as in Fig. 3-1A-B. Motifs A-F, the priming loop (PL), and the VP1-VP2 interaction sites predicted by Estrozi et al. (EC1, EC2, and EC3) are outlined in boxes. The active site is indicated with asterisks. Amino acid positions are listed. Dashes indicate gaps in the protein sequence, light gray shading indicates conservation of amino acid identity, and black shading represents variation in amino acid identity. Genotypes are listed on the left and lineages are defined by brackets. R20 (lineage I) is positioned at the bottom of the alignment due to its amino acid similarities with lineage II genotypes (R4, R6, and R14). Percent identity is shown as a colorized graph above the alignment; green bars represent residues with a high degree of intergenotypic conservation, red bars represent residues with little intergenotypic conservation, and yellow bars are intermediary. The enlarged alignment showing amino acid residues is available in Appendix B.

Fig. 3-5. VP2 genotype consensus amino acid sequence alignment. The schematic shows an amino acid sequence alignment of RVA VP2 consensus sequences for genotypes C1-C15 and C17-C18. Dashes indicate gaps in the protein sequence, light gray shading indicates conservation of amino acid identity, and black shading represents variation in amino acid identity. Genotypes are listed on the left and lineages are defined by brackets. Percent identity is shown as a colorized graph above the segment that corresponds to the sequence below; green bars represent residues with a high degree of intergenotypic conservation, red bars represent residues with little intergenotypic conservation, and yellow bars are intermediary. The VP2 domains and subdomains are represented by a line above the sequence and are colored as in Fig. 3-1D, and the VP1-VP2 interaction sites predicted by Estrozi et al. (EC) is outlined in a box. Amino acid positions are indicated. The enlarged alignment showing amino acid residues is available in Appendix C.

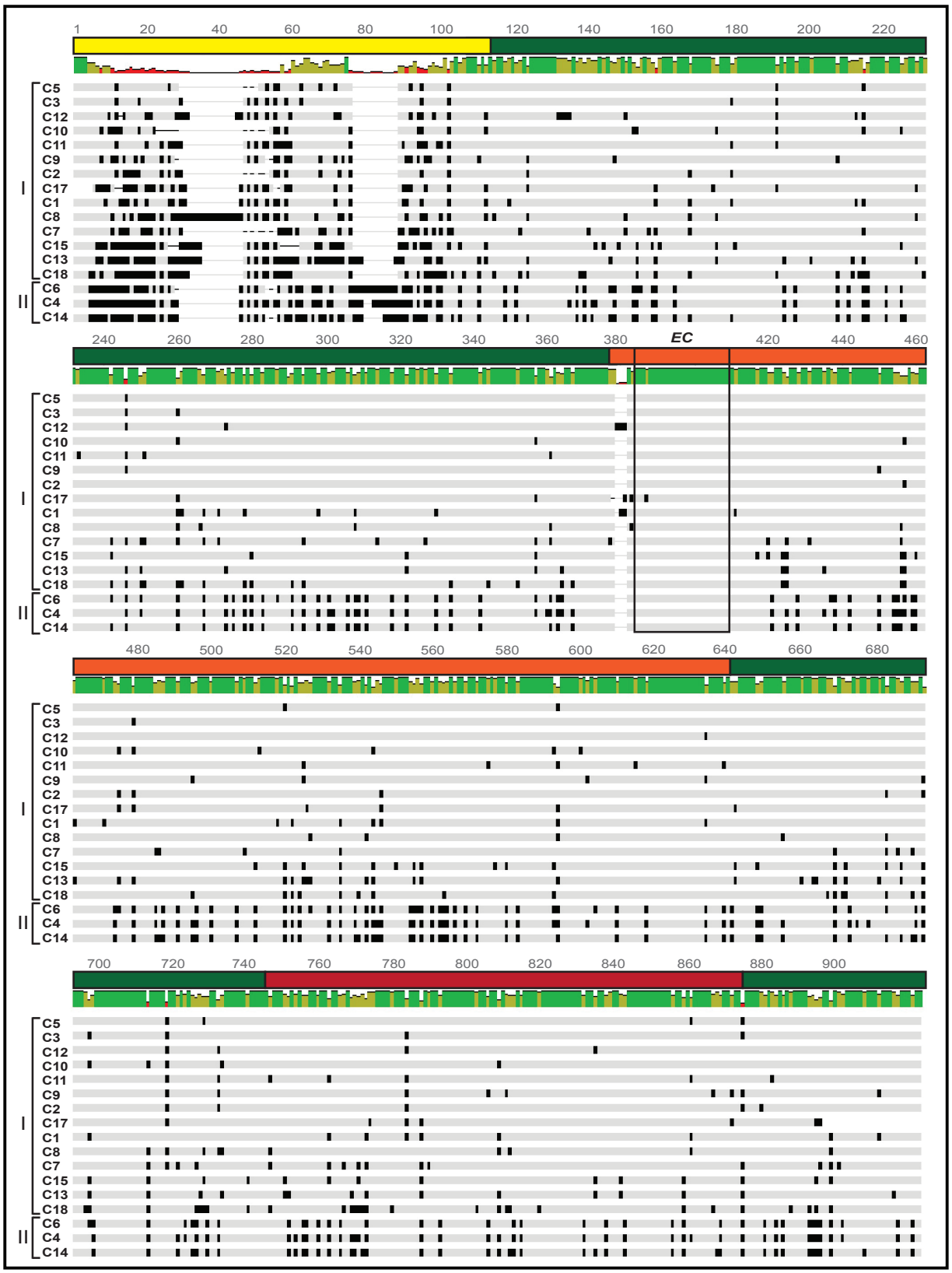


Fig. 3-5. VP2 genotype consensus amino acid sequence alignment. Legend on previous page.

alignment revealed that it has more similarities to avian RVA lineage II proteins (**Fig. 3-4 and Appendix B**). This genotype is clearly an outlier that has unique residues at otherwise conserved positions. For the rest of the genotypes, we found that two of the three sites on VP1 that are predicted to engage VP2 in the context of the DLP (i.e., Estrozi contacts; ECs) were highly conserved, while another showed a greater degree of variation, especially when comparing between lineage I vs. II VP1 proteins. More specifically, the VP2 contact site on VP1 at residues 267-270 (EC1) showed only a single amino acid change when comparing lineage I vs. II proteins, and the site at VP1 residues 1025-1028 (EC3) was 100% conserved (**Fig. 3-4 and Appendix B**). However, the VP2 contact site spanning VP1 CTD residues 971-983 (EC2) varied more drastically, showing as low as 40% sequence identity between some genotypes (e.g., R1 vs. R4) (**Fig. 3-4 and Appendix B**).

Not surprisingly, the catalytic center of the polymerase (motifs A-F and priming loop) showed a higher degree of sequence conservation than did the NTD and CTD (**Fig. 3-4 and Appendix B**). Residues directly implicated in catalysis, including active site aspartates of motif C (SA11 residue D631 and D632), were 100% conserved among the VP1 genotypes (183). However, motif D, which is implicated in polymerase fidelity, was highly divergent when comparing among all the VP1 genotypes and there were several distinguishing residues between lineage I v. lineage II proteins (**Fig. 3-4 and Appendix B**). The R20 VP1 genotype was also an exception to the conservation seen for motifs, showing variable residues at what are otherwise exclusively-conserved positions (**Appendix B**). Specifically, motif B contains a number of residues that anchor the 3' terminus of the RNA template inside the enzyme (G592, E593, K594, and K597 in SA11) or aid in NTP selection and stabilization at the active site (G592, E593, K594, and K597 in SA11); these key residues all differed for the R20 VP1 genotype (**Appendix**

B). Moreover, motif F includes a conserved arginine residue that is positioned to interact with nucleosides in the active site (residue R452 in SA11), which also differed for R20 VP1 (**Appendix B**).

The VP2 genotype consensus alignment revealed a generally more conserved protein compared to VP1, with the dramatic exception of the NTD (**Fig. 3-5 and Appendix C**). More specifically, NTD residues 10-105 showed 28-89% intergenotypic amino acid pairwise sequence distance values and several large insertions/deletions (**Fig. 3-5 and Appendix C**). The VP2 NTD is predicted to engage VP1 in the context of the DLP, but the precise points of contact are unknown because the region is not resolved in the structure (55). Sequence variation in the VP2 principal scaffold domain was less than that seen with the NTD, and it was most dramatic between lineage I v. II core shell proteins (**Fig. 3-5 and Appendix C**). The region of the VP2 apical subdomain that contacts VP1 in the DLP structure (i.e., EC, residues 377-403 of strain UK) is highly conserved among all genotypes. For genotypes C1, C12, and C17 a variable insertion occurred at position 379, just before the contact site (**Fig. 3-5 and Appendix C**).

***In vitro* dsRNA synthesis and functional compatibility of VP1 and VP2 proteins from various genotypes.** We next sought to determine whether the observed intergenotypic variation correlated with VP1 and VP2 protein function. Specifically, we sought to test which of the VP1/VP2 genotypes could functionally substitute for each other in an *in vitro* dsRNA synthesis assay. Proteins that were able to functionally substitute for each other *in vitro* were hypothesized to share sequence features in regions important for VP1-VP2 interaction(s). In contrast, proteins that lacked cross-functionality were expected to be divergent in their sequences. For this analysis, we focused on strains that represent mammalian RVA lineage I

[strain SA11 (R2/C5), strain DS-1 (R2/C2), strain Wa (R1/C1), and strain ETD (R7/C7)], as well as a single strain representing avian RVA lineage II [strain PO-13 (R4 and C4)].

The rVP1 proteins of strains SA11 (R2), DS-1 (R2), Wa (R1), ETD (R7), and PO-13 (R4) were expressed in insect cells as 6X histidine-fusions, purified from the soluble fraction of cell lysates using metal affinity resin, and visualized following SDS-polyacrylamide gel electrophoresis (PAGE) and Coomassie Blue staining (**Fig. 3-6A**). In general, the expression and solubility profiles of the rVP1 proteins were similar, with the exception of strain PO-13 (R4) rVP1 for which there was a lower relative recovery after affinity purification (data not shown). The rVP2 proteins of strains SA11 (C5), DS-1 (C2), Wa (C1), ETD (C7), and PO-13 (C4) were expressed in insect cells as untagged proteins, purified from the cell lysates using different centrifugation, and also visualized following SDS-PAGE and Coomassie Blue staining (**Fig. 3-6B**). The expression and recovery of intact rVP2 was generally similar for all strains, with some exceptions. For example, DS-1 (C2) rVP2 showed lower expression levels compared to the other 4 core shell proteins and had a banding pattern in gels that was consistent with NTD cleavage (46, 198) (**Fig. 3-6B**). As seen in previous studies, additional protein bands migrating between ~35-75-kDa in size were observed at varying concentrations in the rVP2 preparations and may represent co-purifying contaminants (119). Reasons underlying the observed variation in levels of co-purifying proteins and cleavage products in rVP2 protein preparations is not currently well understood, albeit not unprecedented (119, 183).

First, to test whether cognate rVP1 and rVP2 protein sets were functional for *in vitro* dsRNA synthesis, they were incubated together along with an RVA +RNA template, divalent cations, NTPs, and trace amounts of ^{32}P-UTP at 37°C for 3 hours. The ^{32}P-labeled dsRNA products of the reactions were resolved by SDS-PAGE and visualized using a phosphorimager

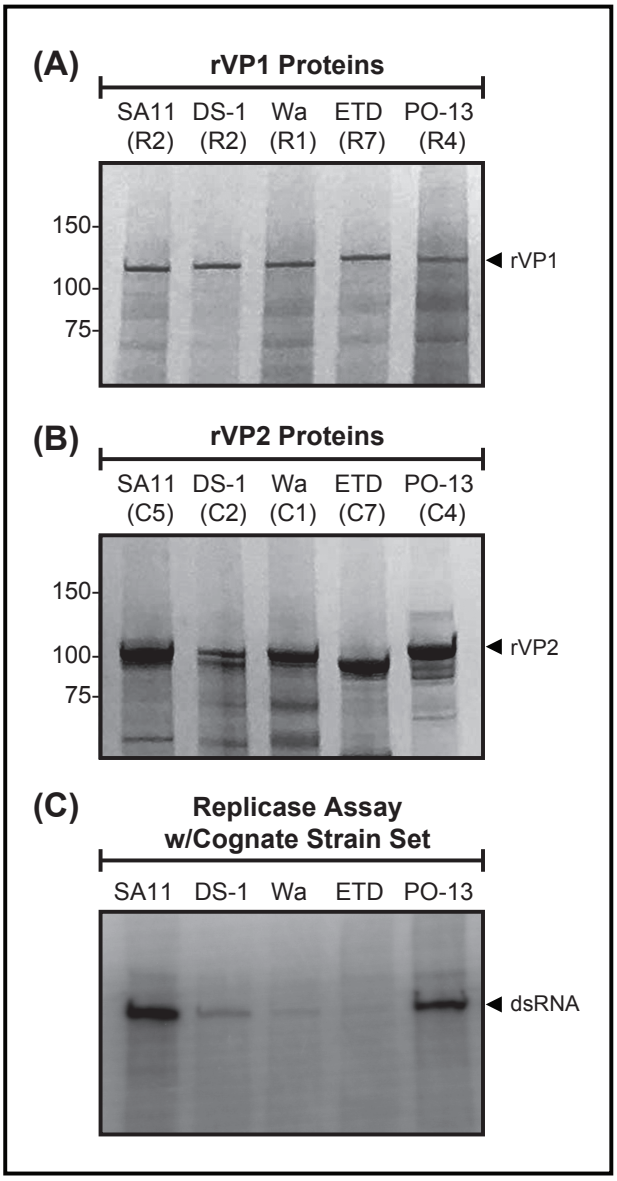


Fig. 3-6. In vitro dsRNA synthesis by rVP1 and rVP2 proteins of several genotypes. Approximately 2 pmol purified rVP1 (A) or 8 pmol purified rVP2 (B) were electrophoresed in 4-15% SDS-polyacrylamide gels and visualized by Coomassie Blue stain. Molecular mass (in kilodaltons) is shown to the left of each gel. Strain name and genotype of the recombinant proteins are specified above the corresponding lane. (C) In vitro dsRNA synthesis assays were performed using 2 pmol of rVP1, 8 pmol of rVP2, and 16 pmol of an RVA +RNA template. Radiolabeled dsRNA products made by rVP1/rVP2 proteins of SA11 (R2/C5), DS-1 (R2/C2), Wa (R1/C1), ETD (R7/C7), and PO-13 (R4/C4) were resolved in 4-15% SDS-polyacrylamide gels and detected with a phosphorimager. Experiments were repeated three times with four distinct protein batches and representative images are shown.

(**Fig. 3-6C**). The results showed strong dsRNA product bands for the reactions containing cognate rVP1 and rVP2 proteins of strains SA11 (R2/C5) and PO-13 (R4/C4), suggesting that both of these polymerases and their core shell proteins were properly folded and functional. In contrast, faint dsRNA product bands were detected for reactions containing rVP1 and rVP2 proteins of strains DS-1 (R2/C2) and Wa (R1/C1); no reproducibly detectable products were observed for reactions containing strain ETD (R7/C7) proteins (**Fig. 3-6C**). These latter results suggest that one or both of the protein binding partners (i.e., rVP1 and/or rVP2) for strains Wa, DS-1, or ETD were not functional *in vitro* under the assay conditions.

To determine which individual rVP1 and/or rVP2 proteins were functional *in vitro* and which ones could substitute for each other, we next performed 'mix-n-match' reactions (**Fig. 3-7**). More specifically, each rVP1 was assayed for the capacity to support *in vitro* dsRNA synthesis in the presence of each rVP2. The results showed strong dsRNA product bands for reactions containing rVP1 proteins of the mammalian RVA lineage I genotype proteins from strains SA11 (R2), DS-1 (R2), and Wa (R1) in the presence of strain SA11 (C5) rVP2 (**Fig. 3-7A-C**). Faint, but reproducibly-detectable, dsRNA product bands were also seen for reactions containing these same rVP1 proteins in the presence of strain ETD (C7) rVP2 (**Fig. 3-7A-C**). Thus, these 3 mammalian RVA polymerases [i.e., SA11 (R2), DS-1 (R2), and Wa (R1)] and these 2 mammalian RVA core shell proteins [i.e., SA11 (C5) and ETD (C7)] were functional for *in vitro* dsRNA synthesis and were functionally compatible with each other. In contrast, there were little or no detectable dsRNA products for reactions containing mammalian RVA lineage I ETD (R7) rVP1, suggesting that this polymerase was not functional under the assay conditions for reasons that are not understood at this time (**Fig. 3-7D**). Likewise, little or no dsRNA products were detected in reactions containing lineage I rVP2 proteins from human strains DS-1

(C2) and Wa (C1), even in the presence of functional cognate polymerases [i.e., DS-1 (R2) rVP1 and Wa (R1) rVP1] (**Fig. 3-7A-C**). The reason for the lack of activity with these 2 mammalian RVA core shell proteins is not understood but might be due to proteolysis or contaminant proteins in the preparations (**Fig. 3-6B**). For the reactions containing the avian RVA lineage II

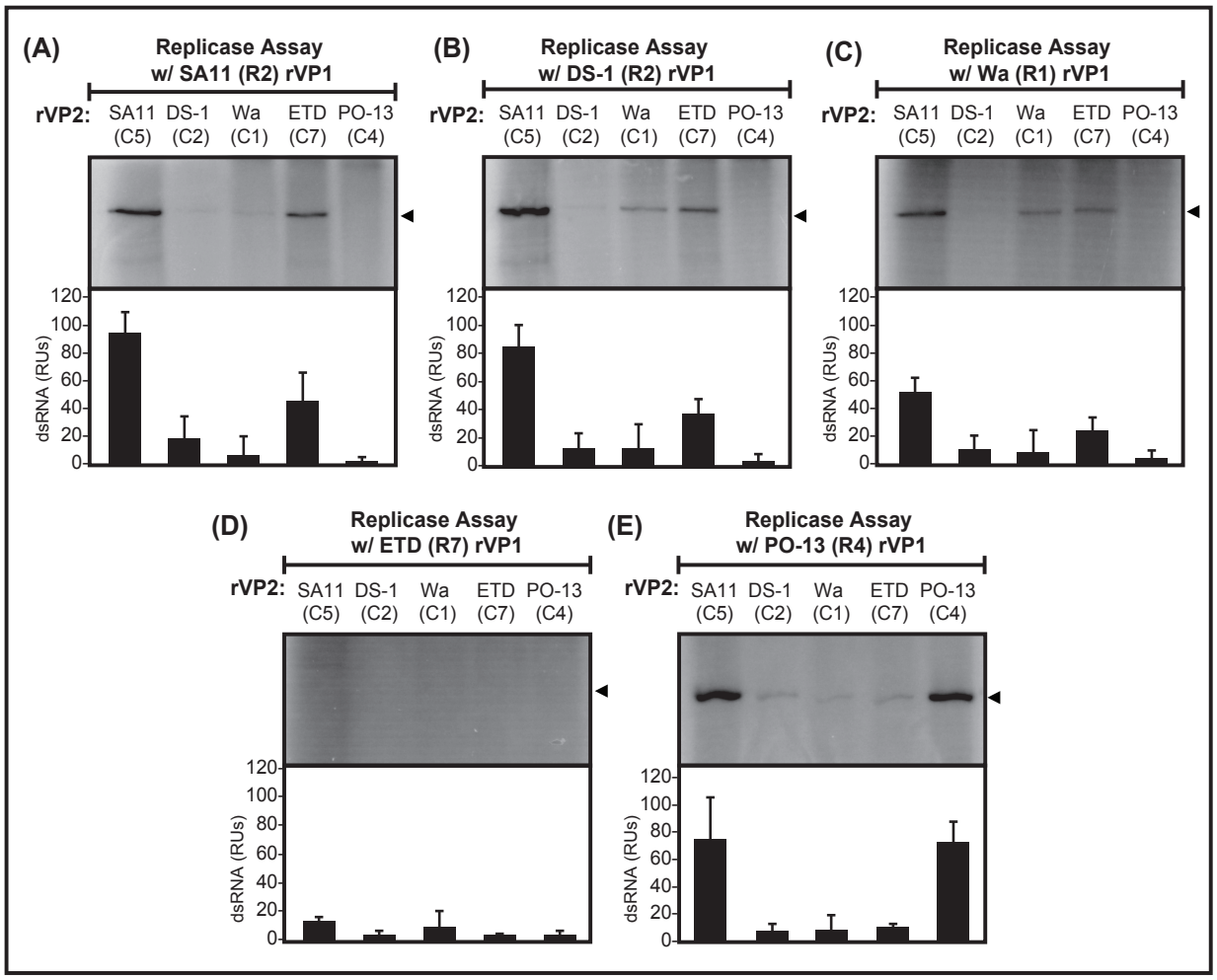


Fig. 3-7. In vitro functional compatibility of rVP1 and rVP2 proteins. In vitro dsRNA synthesis assays were performed using 2 pmol of SA11 (R2) rVP1 (A), DS-1 (R2) rVP1 (B), Wa (R1) rVP1 (C), ETD (R7) rVP1 (D), or PO-13 (R4) rVP1(E) in the presence of 8 pmol of rVP2 protein from strains SA11 (C5), DS-1 (C2), Wa (C1), ETD (C7), and PO-13 (C4) in a 'mix-n-match' format. All reactions contained 16 pmol of RVA +RNA template and were incubated at 37°C for 180 minutes. Radiolabeled dsRNA products (indicated with arrowheads) were resolved in 4-15% SDS-polyacrylamide gels and detected with a phosphorimager. Radiolabeled dsRNA from three independent experiments using at least two independent protein batches was quantified and expressed as relative units (RUs). Averages are shown as bar graphs below each gel, and error bars represent standard deviation from the mean.

PO-13 (R4) rVP1 protein, robust dsRNA product bands were detected in the presence of its own cognate PO-13 (C4) rVP2 and also in the presence of the mammalian SA11 (C5) rVP2 (**Fig. 3-7E**). None of the other mammalian RVA rVP2 proteins, not even the functional ETD (C7) rVP2, were able to support activity with PO-13 (R4) rVP1. Importantly, none of the functional rVP1 proteins [i.e., strains SA11 (R2), DS-1 (R2), Wa (R1), and PO-13 (R4)] produced detectable dsRNA product in the absence of rVP2, supporting the essential nature of the VP1-VP2 interaction during genome replication (data not shown). Together, the results of this 'mix-n-match' experiment raise several questions about the precise determinants of rVP1 activation by rVP2 and how intergenotypic sequence diversity plays a role in functional compatibility.

Mapping regions of rVP1 that are important for activation by rVP2. We next sought to follow up on the observed differences in *in vitro* dsRNA synthesis for the rVP1/rVP2 proteins of strains SA11 (R2/C5) and PO-13 (R4/C4). Specifically, we used the lack of functional compatibility of SA11 (R2) with PO-13 (C4) as a basis for mapping key regions of the polymerase that might be important for recognition by the avian RVA lineage II core shell protein. We first generated a sequence alignment to identify regions of variation between SA11 (R2) and PO-13 (R4) (**Fig. 3-8A**). Then, we plotted these differences onto the atomic structure of SA11 (R2) VP1 to determine their three-dimensional (3D) locations (**Fig. 3-8B-E**). While a total of 257 residues differ between SA11 (R2) and PO-13 (R4) VP1, for this study, we focused on regions of sequence diversity that were surface-exposed and located on the side of VP1 that includes the +RNA exit tunnel. This is the region of VP1 that interacts with VP2 in that context of the DLP (i.e., transcription complex) (55). Since the enzymatic function of VP1 is VP2-dependent for both transcription and genome replication, we posited that this VP1 region might

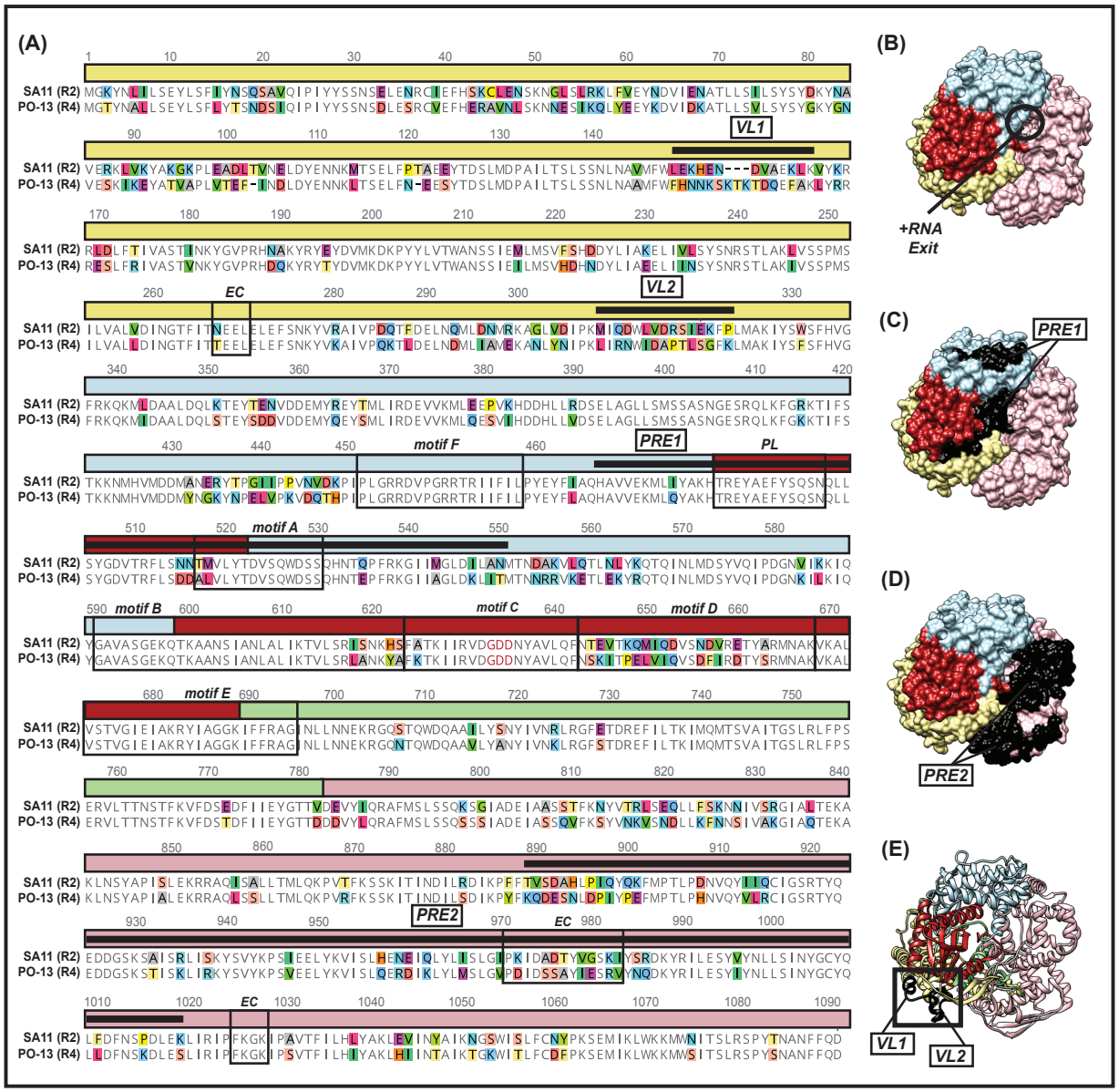


Fig. 3-8. Amino acid differences between SA11 (R2) VP1 and PO-13 (R4) VP1. (A) Primary amino acid sequence alignment of SA11 (R2) VP1 and PO-13 (R4) VP1. Dashes indicate gaps in the protein sequence, colored shading represents variation in amino acid identity. The VP1 domains and subdomains are represented by a line above the sequence and are colored as in Fig. 3-1A-B. Motifs A-F, priming loop (PL), and VP1-VP2 interaction sites (EC1, EC2, and EC3) are identified in boxed regions. Surface-exposed regions surrounding the +RNA exit tunnel (PRE) and two surface-exposed loops (VL) are indicated with thick black lines above the sequence. The active site is indicated with asterisks. Amino acid position is listed above the alignment. (B) Atomic structure of SA11 (R2) rVP1 (PDB accession no. 2R7R) shown in surface representation with the +RNA exit tunnel facing forward and labeled. (C-D) Structure of SA11 (R2) rVP1 shown in the same orientation as panel B with PRE1 (C) or PRE2 (D) residues colored in black and labeled. (E) Structure of SA11 (R2) rVP1 in the same orientation as panels B-D depicted in ribbon representation with VL1 and VL2 residues colored in black and labeled.

also be important for binding/activation by VP2 during *in vitro* dsRNA synthesis. Thus, the regions of VP1 that were targeted for mutagenesis included: (i) two divergent +RNA exit interface regions of the polymerase domain (PRE1, SA11 VP1 residues 474-547) or the VP1 CTD (PRE2, SA11 VP1 residues 886-1016) were exchanged individually or together (**Fig. 3-8A, C-D**) and (ii) two surface-exposed, variable helix-loop regions in the VP1 NTD (VL1 and VL2, SA11 VP1 residues 149-161 and 306-320, respectively) were exchanged individually (**Fig. 3-8A and E**). We generated chimeric rVP1 proteins in which the aforementioned regions of SA11 (R2) sequence were swapped with corresponding sections of PO-13 (R4) sequence. We predicted that the chimeric rVP1 proteins would maintain their capacity to produce dsRNA in reactions containing by SA11 (C5) rVP2 but that they would now also gain functional activity and produce dsRNA in reactions containing PO-13 (C4) rVP2. Molecular dynamics (MD) simulations of the SA11 (R2) atomic structure vs. a PO-13 (R4) homology model revealed that VL1 is the only region predicted to be differentially flexible, with the exception of a modeled flexible loop (SA11 VP1 residues 346-358) that is missing from the SA11 (R2) atomic structure (**Fig. 3-9A**). Residues 155-161 of VL1, in particular, show greater backbone mobility in PO-13 (R4) VP1 vs. SA11 (R2) VP1 as measured by B-factors calculated from the root mean square fluctuation of a residue's alpha-carbons over the course of an MD simulation (**Fig. 3-9B**). The other surface-exposed regions surrounding the +RNA exit tunnel of VP1 (e.g., VL2, PRE1 and PRE2), however, do not display such changes in local flexibility across the different polymerase variants, although they differ in their primary sequence.

The chimeric rVP1 proteins (PRE1, PRE1+2, VL1, and VL2) were expressed alongside wildtype control proteins [SA11 (R2) rVP1 and PO-13 (R4) rVP1] in insect cells as 6X histidine-fusions, purified from the soluble fraction of cell lysates using metal affinity resin, and visualized

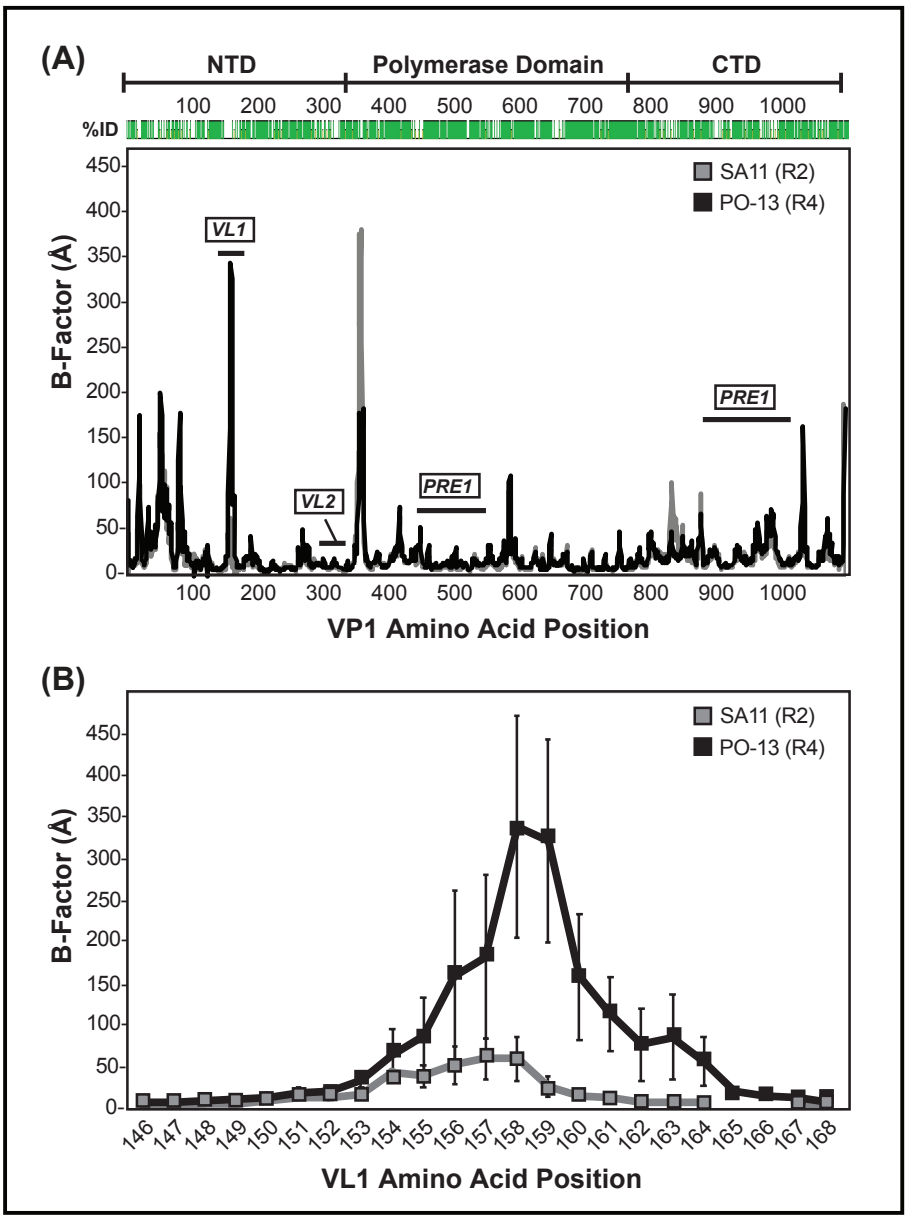


Fig. 3-9. Molecular dynamics simulation of VP1 protein structures. Molecular dynamics simulations were performed using GROMACS v5.1.347 on the atomic structure of the SA11 (R2) VP1 (PDB accession no. 2R7R) or on a homology model of PO-13 (R4) VP1 (PDB file available upon request) as described in McKell et al., 2017. Three trajectories initiated with different random seeds were run for each protein structure. The RMSF of alpha-carbons from each of the three trajectories was calculated. B-factors for each residue were calculated from the RMSF values. Average B-factors are shown for (A) the entire VP1 proteins and (B) the VL1 regions (residues 146-168). In both panels A and B, residue position numbers are shown on the X-axis, SA11 (R2) VP1 B-factors are plotted as gray boxes and PO-13 (R4) VP1 B-factors are plotted as black boxes. Error bars represent standard deviation from the mean following three independent simulations. In panel A, the % amino acid sequence identity values are shown in graphical representation and the domains and regions of interest according to Fig. 3-8 for VP1 are indicated.

following SDS-PAGE and Coomassie Blue staining (**Fig. 3-10A-B**). The expression and solubility profiles of the chimeric rVP1 proteins were indistinguishable from that of SA11 (R2) rVP1, suggesting that mutant polymerases folded well enough so as to remain soluble and expose their epitope tags. In contrast, we attempted to express/purify additional chimeric proteins that included a swap of motif D (residues 642-668) or several different regions of the VP1 NTD; these proteins were insoluble, suggesting they were grossly misfolded (data not shown).

The purified chimeric mutant rVP1 (PRE1, PRE1+2, VL1, and VL2) and wildtype counterpart rVP1 were then assayed for the capacity to synthesize dsRNA *in vitro* in the presence of either SA11 (C5) rVP2 or PO-13 (C4) rVP2 (**Fig. 3-10C-D**). The results with SA11 (C5) rVP2 show that PRE1 rVP1 and VL1 rVP1 produced detectable, but substantially diminished, dsRNA products as compared to the control proteins (**Fig. 3-10C**). Thus, the 10 PO-13 (R4)-specific amino acid changes that were introduced into the SA11 (R2) rVP1 protein (i.e., to create PRE1 rVP1) caused a loss-of-function phenotype in the polymerase. Even more, PRE1+2 showed no detectable dsRNA products, suggesting that the 39 additional PO-13 (R4)-specific mutations to account for PRE2 [i.e., 49 total changes in SA11 (R2) rVP1] completely ablated enzyme activity. VL1 also showed reduced activity compared to the controls, suggesting the 13 PO-13 (R4)-specific residues in this surface loop also caused a loss-of-function phenotype for SA11 (R2) rVP1. In contrast, VL2 rVP1 produced dsRNA products at levels comparable to both control proteins, suggesting that this chimera was fully functional and that the 12 PO-13 (R4)-specific amino acid changes had no effect on SA11 (R2) rVP1 function. Unfortunately, none of chimeric proteins, not even VL2, produced dsRNA products in reactions containing

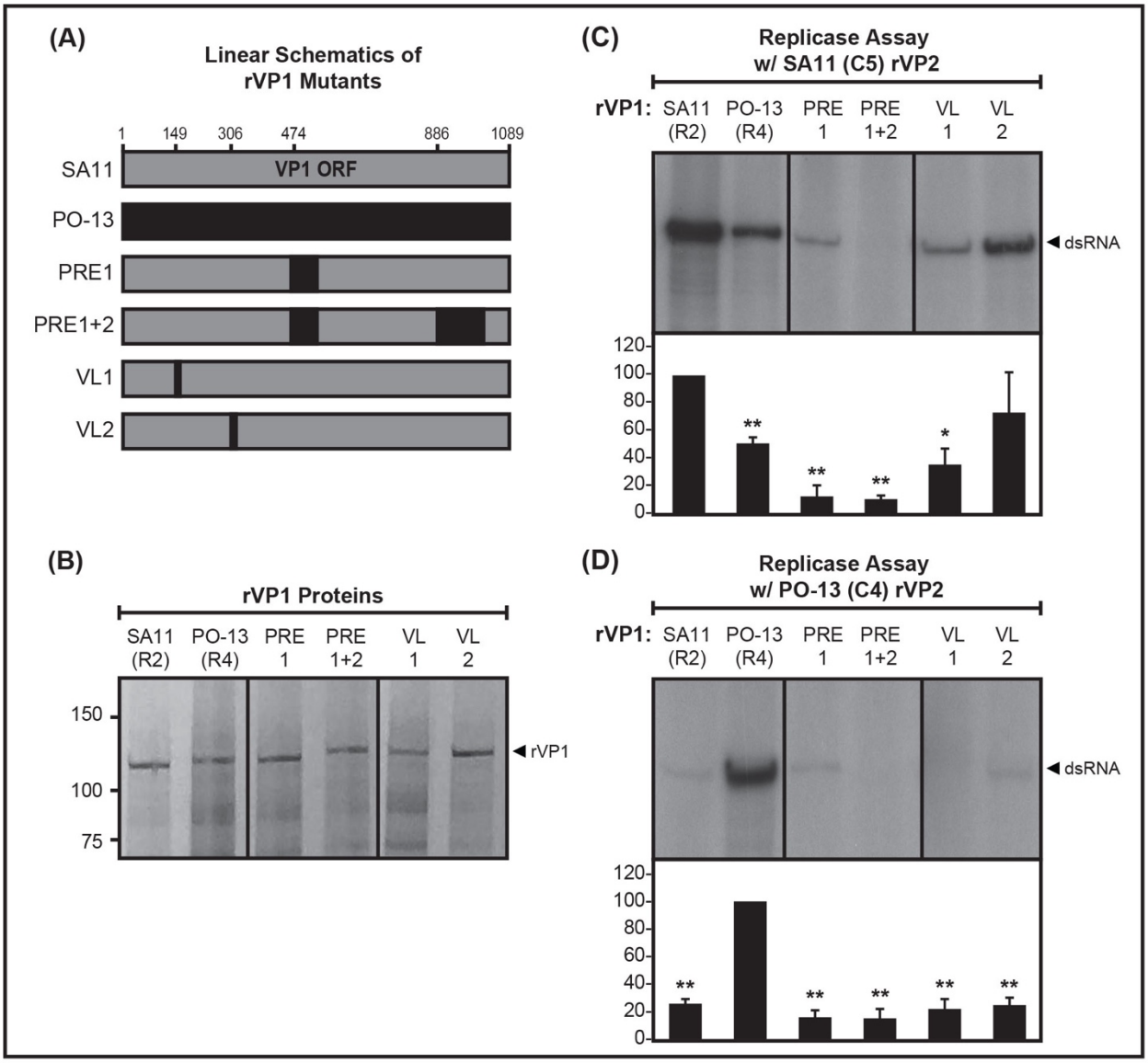


Fig. 3-10. In vitro dsRNA synthesis by chimeric rVP1 proteins. (A) Linear schematics of recombinant VP1 proteins are shown as boxes. For all genes, SA11 sequence is shown in gray and PO-13 sequence is shown in black. Amino acid numbers are listed above the schematic and general locations of mutated regions are shown. (B) Approximately 2 pmol purified SA11 (R2) rVP1, PO-13 (R4) rVP1, PRE1 rVP1, PRE1+2 rVP1, VL1 rVP1, or VL2 rVP1 were electrophoresed in a 4-15% SDS-polyacrylamide gel and visualized by Coomassie Blue stain. Molecular mass (in kilodaltons) is shown to the left. (C-D) Radiolabeled dsRNA synthesis products synthesized by 2 pmol of each rVP1 presence of 8 pmol of rVP2 from strain SA11 (C5) rVP2 (C) or PO-13 (C4) rVP2 (D). All reactions contained 16 pmol of an RVA +RNA template and were incubated at 37°C for 180 minutes. Radiolabeled dsRNA products were resolved in 4-15% SDS-polyacrylamide gels and visualized using either a phosphorimager or autoradiography. Radiolabeled dsRNA from three independent experiments using at least three protein batches was quantified and expressed as relative units (RUs). Averages are shown as bar graphs below each gel, and error bars represent standard deviation from the mean. A single asterisk indicates a P value of <0.05 and two asterisks indicates $P < 0.005$.

PO-13 (R4) rVP2 (**Fig. 3-10D**), suggesting that the 12 PO-13 (R4)-specific amino acids were not sufficient to swap the core shell specificity of the polymerase.

Mapping regions of rVP2 that are important for rVP1 activation. We next turned our attention toward the differences in sequences and *in vitro* functions of the mammalian SA11 (C5) and avian PO-13 (C4) core shell proteins. Specifically, we sought to determine which core shell protein regions are important for supporting *in vitro* dsRNA synthesis from either a mammalian SA11 (R2) polymerase or avian RVA (R4) polymerase. To understand the extent of sequence variation between the SA11 (C5) and PO-13 (C4) VP2 core shell proteins, we first created an amino acid sequence alignment (**Fig. 3-11**). Altogether, the proteins differ by 219 residues, with 64 of them being changes in the NTD (residues 1-117 in the alignment) and 155 being changes in the principal scaffold domain (residues 118-897 in the alignment). Based upon previous work with RVA and RVC polymerases and core shell proteins, we hypothesized that (i) the specificity of PO-13 (C4) rVP2 for its cognate PO-13 (R4) rVP1 would reside in the principal scaffold domain and that (ii) divergent SA11 (C5) vs. PO-13 (C4) NTDs would not be able to functionally substitute for each other (119). To test this hypothesis, we engineered a chimeric rVP2 containing the SA11 (C5) NTD (residues 1-102) fused to the PO-13 (C4) principal scaffold domain (residues 118-897) (i.e., SA:PO rVP2) or the reverse chimera (i.e., PO:SA rVP2) (**Fig. 3-12A**). The chimeric rVP2 proteins were expressed alongside a wildtype PO-13 (C4) rVP2 control in insect cells, purified from the cell lysates by centrifugation, and visualized following SDS-PAGE and Coomassie Blue staining (**Fig. 3-12B**). Both the SA:PO and PO:SA core shell proteins expressed and purified in a manner indistinguishable from the wildtype PO-13 (R4) rVP2 control. However, additional bands were detected on the gel for SA:PO rVP2 and may represent proteolysis products.

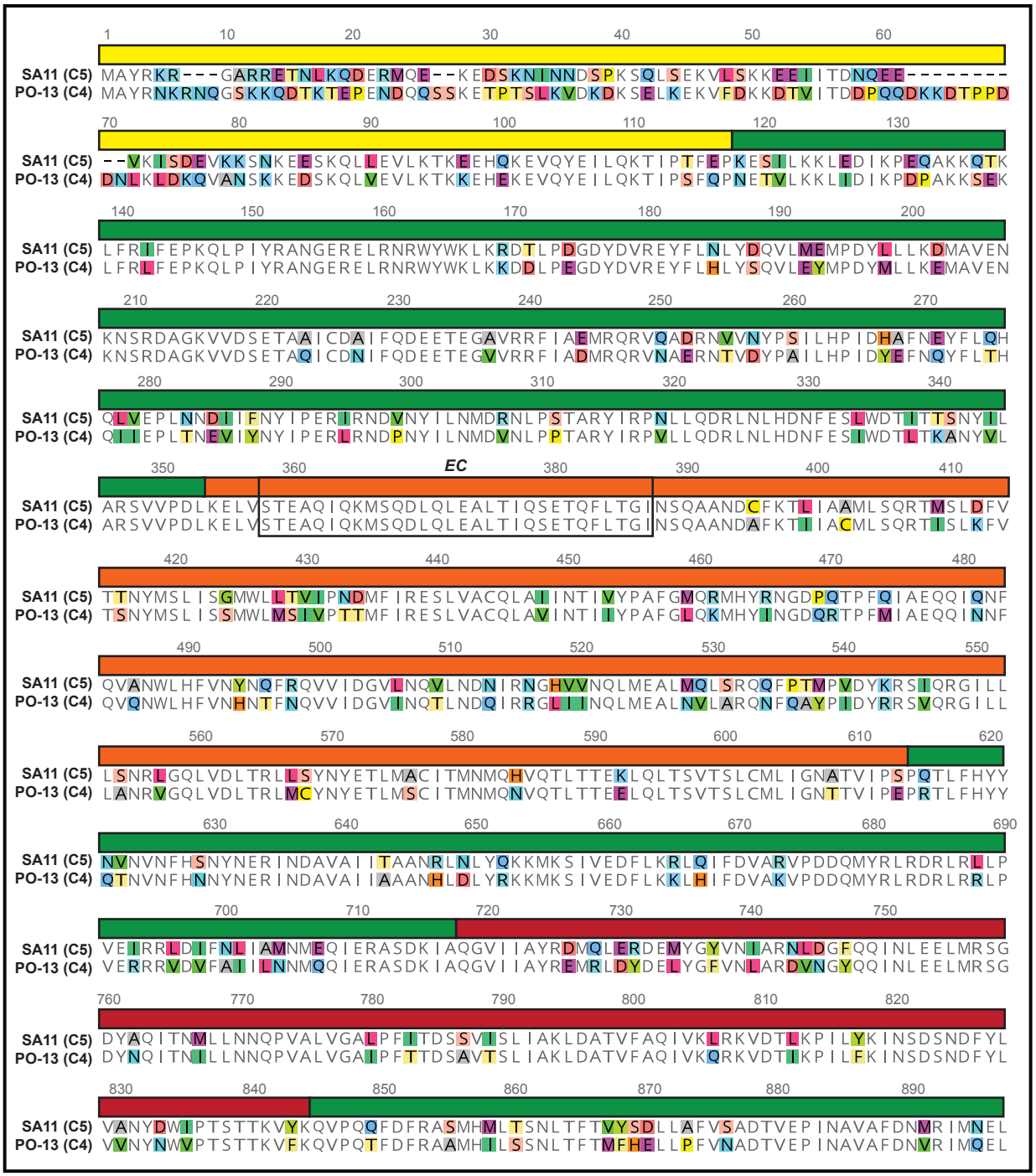


Fig. 3-11. Amino acid differences between SA11 (C5) VP2 and PO-13 (C4) VP2. (A) Primary amino acid sequence alignment of SA11 (C5) VP2 and PO-13 (C4) VP2 is shown. Dashes indicate gaps in the protein sequence, colored shading represents variation in amino acid identity. The VP2 domains and subdomains are represented by a line above the sequence and are colored as in Fig. 3-1D. The VP1-VP2 interaction site predicted by Estrozi et al. (EC) is outlined with a box. Amino acid position is indicated above the alignment.

The purified, chimeric mutant rVP2 (SA:PO and PO:SA) and wildtype PO-13 (C4) rVP2 were then assayed for the capacity to support dsRNA synthesis *in vitro* by either the SA11 (R2) rVP1 or PO-13 (R4) rVP1 (**Fig. 3-12C-D**). The results with SA11 (R2) rVP1 show that robust dsRNA products were made in reactions containing PO:SA rVP2, but not SA:PO rVP2, suggesting that the SA11 (C5) NTD is not sufficient to mediate activation specificity and consistent with the hypothesis that residues critical for the specific recognition of this mammalian RVA polymerase reside in the principal scaffold domain (**Fig. 3-12C**) The results with PO-13 (R4) rVP1 show that robust dsRNA products were made in reactions containing PO:SA rVP2 and that lower levels of dsRNA products were made in the presence of SA:PO rVP2 (**Fig. 3-12D**). The loss-of-function phenotype for SA:PO rVP2 relative to PO-13 (C4) rVP2 indicates that the SA11 (C5) NTD cannot functionally substitute for that of PO-13 (C4) rVP2. Thus, the avian RVA polymerase prefers an avian PO-13 (C4) rVP2 NTD, indicating that this unstructured and highly-variable region of the core shell protein (and the 64 amino acid differences wherein) might facilitate interaction of VP1 with the principal scaffold domain.

Discussion

RVAs are widespread in nature, infecting numerous mammalian and avian hosts and causing severe gastroenteritis in human children (6, 176). To classify the diversity and genetic relationships among RVA strains, a system was developed that assigns a specific genotype to each of the 11 genome segments according to its sequence and established nucleotide percent identity cut-off values (34) (**Table 3-1**). To date, 20-51 different genotypes have been described for RVA genes, and these genotypes can combine in a variety of different ways due to segment reassortment (34, 109). Some genotype combinations seem to be favored in certain host species,

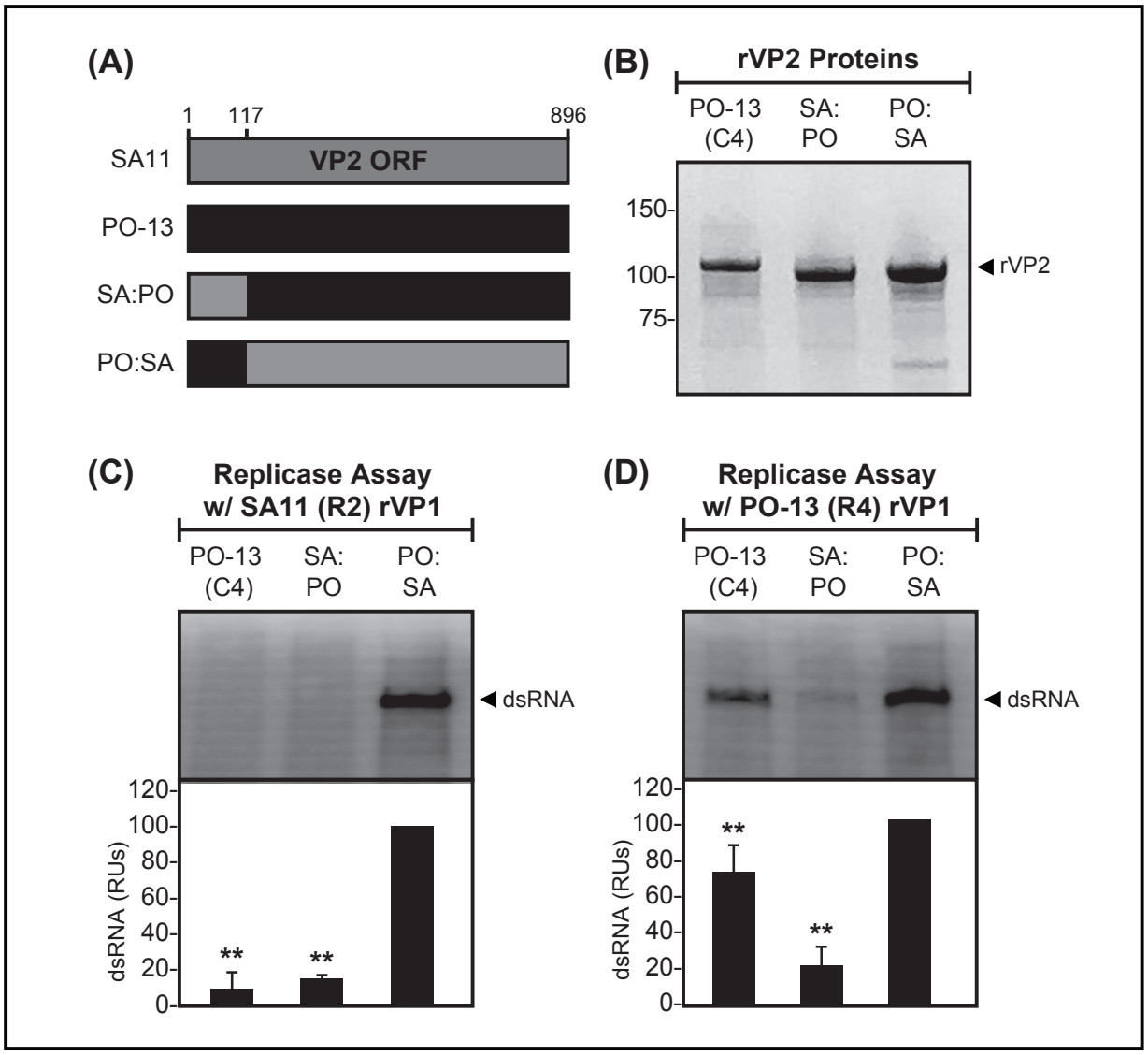


Fig. 3-12. In vitro dsRNA synthesis by chimeric rVP2 proteins. (A) Linear schematics of recombinant VP2 proteins are shown as boxes. For all genes, SA11 sequence is shown in gray and PO-13 sequence is shown in black. Amino acid numbers are listed above the schematic and general locations of mutated regions are shown. (B) Approximately 8 pmol purified PO-13 (C4) rVP2, SA:PO rVP2, or PO:SA rVP2 were electrophoresed in a 4-15% SDS-polyacrylamide gel and visualized by Coomassie Blue stain. Molecular mass (in kilodaltons) is shown to the left. Radiolabeled dsRNA was synthesized by 2 pmol of either SA11 (R2) rVP1 (C) or PO-13 (R4) rVP1 (D) in the presence of 8 pmol of each rVP2. All reactions contained 16 pmol of an RVA +RNA template and were incubated at 37°C for 180 minutes. Radiolabeled dsRNA products were resolved in 4-15% SDS-polyacrylamide gels and visualized using a phosphorimager. Radiolabeled dsRNA from three independent experiments using at least two protein batches was quantified and expressed as relative units (RUs). Averages are shown as bar graphs below each gel, and error bars represent standard deviation from the mean. Double asterisks indicate P values of <0.005.

perhaps because the encoded proteins function better when they are kept together and not unlinked by reassortment (109, 193). However, the extent to which nucleotide-level intergenotypic diversity is reflected in the sequences, structures, or functions of the encoded proteins has not yet been fully described for RVAs. For instance, 22 different genotypes have been defined for the gene encoding the VP1 polymerase based upon an 83% nucleotide identity cut-off value. Thus, VP1 genes must differ by at least 17% in their nucleotide sequences to be considered distinct genotypes. Due to redundancy in the genetic code, however, ORF nucleotide sequences can differ by ~30% and still translate the same amino acid sequence. Therefore, it is theoretically possible that some VP1 genotypes encode identical or nearly identical polymerase proteins. The same idea holds true for the 20 different genotypes that are defined for the gene encoding the VP2 core shell protein based upon an 84% nucleotide cut-off value. For that reason, one goal of this study was to determine the level of divergence among RVA VP1 and VP2 proteins that are encoded by the currently-recognized genotypes.

We employed several complementary approaches to study the intergenotypic diversity and relationships of the VP1 and VP2 at the protein level. MDS was used to visualize the relationships among VP1 and VP2 genotypes based upon their pairwise amino acid sequence distances (**Fig. 3-2B and 3-3B**). For both VP1 and VP2 proteins, the results showed that the avian RVA lineage II genotypes (i.e., R4, R6, R14, C4, C6, and C14) were distinct from the mammalian RVA lineage I genotypes. This result was consistent with our ORF nucleotide sequence-based phylogenetic analyses, which showed strongly-supported branch separation for lineage I vs. II genes, as well as with our identification of numerous lineage-specific amino acid signatures for VP1 and VP2 (**Fig. 3-2A and 3-3A**). Thus, the data suggest that the polymerase and core shell proteins of avian RVAs are quite different from those of mammalian RVAs at

both the nucleotide and amino acid sequence level. However, the within lineage results were not as straight-forward, especially regarding to the similarities among the mammalian RVA lineage I proteins. For example, many of the mammalian RVA VP1 genotypes (i.e., R2, R3, R5, R8-12, and R18) and VP2 genotypes (i.e., C2, C3, C5, C8-13, C15, and C17) could not be differentiated from each other based upon pairwise amino acid sequence distances in the MDS analysis nor did they show any defining signature sequences (**Fig. 3-2 and 3-3**). Still, other mammalian RVA genotypes (e.g., R1 and C1) showed stronger definition by both by MDS and branch-specific amino acid analysis (**Fig. 3-2 and 3-3**). One caveat to our analysis is that we had limited sequence information for several of the more recently-identified genotypes (i.e., R8-R12, R14-18, C8-14, C17-18). It is possible that defining features will be more apparent once additional sequences are available for these genotypes. It is also important to note that 6 genotypes (R17, R21, R22, C16, C19, and C20) were not included in our analysis because their sequences were not available in the database when this study began. Future studies in our laboratories will continue to assess VP1 and VP2 sequences as new genotype sequences become available, with the overall goal of understanding the impact of this naturally-occurring sequence diversity on VP1/VP2 structure, function, and interactions. Even more, our long-term goal is to extend this analysis to the other 9 viral proteins, thereby providing a comprehensive understanding of protein-level RVA diversity in nature.

As an extension of our intergenotypic bioinformatic analyses, we also wanted to identify which regions on the VP1 and VP2 proteins tolerated amino acid changes vs. which regions remained largely invariant, so as to inform an understanding of functional and interactional domains. To do this, we created genotype consensus amino acid sequence alignments and interpreted the conserved/variable regions in light of known protein domains, subdomains, and

motifs (**Fig. 3-4, 3-5 and Appendices B-C**). Somewhat to our surprise, the variation in VP1 was fairly distributed throughout the protein sequence and was seen even in some of the catalytic motifs of the polymerase domain (**Fig. 3-4 and Appendix B**). Motif D, which is implicated in polymerase fidelity, was highly variant in sequence among all of the genotypes (**Fig. 3-4 and Appendix B**). Also, R20 showed several deviations from consensus for motifs B and F, implying that more evolutionary plasticity can occur within the catalytic core than what was previously thought (183) (**Fig. 3-4 and Appendix B**). However, there is only a single representative of the R20 genotype, which was derived from a metagenomic analysis of bat feces; therefore, it is possible that these R20-specific changes are sequencing errors (189). In general, the VP2 contact points on VP1 that are predicted by Estrozi et al. in the DLP structure were conserved across all of the genotypes, with the exception of the site spanning CTD residues 971-983 (i.e., EC2) (55) (**Fig. 3-4 and Appendix B**). This region showed several amino acid changes, especially when comparing the mammalian RVA lineage I vs. avian RVA lineage II VP1 proteins (**Appendix B**).

The VP2 alignment was generally more conserved than that of VP1 with the dramatic exception of the NTD (residues ~1-100), which showed minimal sequence similarities among the genotypes (**Figs. 3-4 and 3-5**). This result confirms and extends previous studies and suggests that the VP2 NTD can tolerate a vast amount of variation, including large insertions and deletions in its sequence (119, 199). The VP2 NTD is non-essential for formation of the T=1 core shell, but it does play an important role in VP1 engagement and activation (46, 119, 183, 198, 200). However, this region does not seem to mediate sequence-specific VP1 interactions; instead, it may play a structural role in supporting VP1 contacts with the principal scaffold domain of VP2 (119). More specifically, the results using NTD chimeras in this study (**Fig. 3-**

12) and in a report by McDonald et al. show that strain-specific VP1 activation correlates with the VP2 principal scaffold domain, not the NTD (119). Interestingly, the VP1 contact points within principal scaffold domain that are predicted from the DLP structure (EC; residues 377-403 of strain UK) are >99% conserved among the genotypes (55) (**Appendix C**). Indeed, if this was the key site of VP1 binding for enzymatic activation during dsRNA synthesis, one would expect that all of the different VP2 genotype proteins would functionally substitute for each other (see also below).

Having observed the regions of amino acid sequence variation among the VP1 and VP2 genotypes, our next goal in this study was to determine if/how this variation correlated with *in vitro* functional interactions between the proteins. More specifically, we employed an established *in vitro* assay in which rVP1 performs minus-strand RNA synthesis on a +RNA template (to produce dsRNA) only when it is engaged/activated by rVP2. We reasoned that if different genotypes could functionally substitute for each other *in vitro*, they must show conserved sequences at their binding interfaces. In contrast, those genotype that do not substitute for each other would show sequence variations at binding interfaces. For this analysis, we chose to use the rVP1 and rVP2 proteins of five different RVA strains (SA11, DS-1, Wa, ETD, and PO-13) that include representatives of mammalian RVA lineage I genotypes (R1, R2, R7, C1, C2, C5, and C7), as well as a single strain representing avian RVA lineage II genotypes (R4 and C4). We tested all rVP1 genotypes for the capacity to synthesize dsRNA in the presence of all rVP2 genotypes (**Fig. 3-7**). Unfortunately, the ETD (R7) polymerase and the DS-1 (C2) and Wa (C1) VP2 proteins were not fully functional *in vitro* under the tested conditions. Still, the results from the 'mix-n-match' reactions with the 4 functional polymerases [SA11 (R2), DS-1 (R2), Wa

(R1), and PO-13 R4)] and the 3 functional core shell proteins [SA11 (C5), ETD (C7) and PO-13 (C4)] yielded several important observations.

The first observation is that lineage I rVP1s seem to prefer lineage I rVP2s. More specifically, we observed that the mammalian RVA lineage I rVP1s [SA11 (R2), DS-1 (R2) and Wa (R1)] were only activated by mammalian RVA lineage I rVP2 proteins [SA11 (C5) and ETD (C7)] and NOT the avian RVA lineage II PO-13 (R4) genotype (**Fig. 3-7**). This result supports the notion that lineage I rVP1s may have evolved interaction specificity for lineage I rVP2s and lost the capacity to interact with lineage II rVP2. Still, it will be important to test this notion in future studies using other lineage II rVP2 genotypes (e.g., C6 and C14). If the observation holds true, however, the lineage-specific amino acid signatures that we identified via our analysis will be useful in mapping the activation specificity determinants (**Figs. 3-1A and 3-2A**).

The second observation is that SA11 (C5) rVP2 seems to be a 'promiscuous' core shell protein. In particular, the result that SA11 (C5) rVP2 robustly activated *both* the mammalian RVA rVP1 proteins [SA11 (R2), DS-1 (R2), and Wa (R1)] as well as the avian PO-13 (R4) rVP1 suggests that this is a broadly-functional core shell protein that can engage many divergent polymerases. At this time, we do not know why SA11 (C5) rVP2 is so active *in vitro* or whether this is a general characteristic of C5 genotype proteins. Nevertheless, this result indicates that SA11 (C5) rVP2 might engage the polymerases via a conserved region(s) and in a manner that is different from how PO-13 (C4) rVP2 engages its cognate polymerase (see below).

The third observation is that PO-13 (C4) rVP2 seems to be 'faithful' to its own polymerase. For instance, the result that PO-13 (C4) rVP2 specifically activated the avian PO-13 (R4) polymerase, but *not* the mammalian RVA polymerases [SA11 (R2), DS-1 (R2), or Wa (R1)] suggests that this core shell protein is specific to its cognate polymerase. In the future, it

will be interesting to assay the additional lineage II genotype proteins (R6, R14, C6, and C14) to see if this result is simple a reflection of avian RVA lineage II specificity. Still, the current result indicates that residues differing between PO-13 (R4) and SA11 (R2) are important for binding/activation by the avian RVA core shell protein. We were especially intrigued by these results and wanted to use the differences between these polymerase proteins as a platform to map VP1-VP2 interaction regions that are critical for *in vitro* dsRNA synthesis. Specifically, we attempted to engineer an SA11 (R2) rVP1 protein that contained enough PO-13 (R4) residues that it would gain the capacity to be activated by PO-13 (C4) rVP2 (**Figs. 3-8 and 3-10**). We focused on swapping out surface-exposed residues that lie at the +RNA exit interface of VP1, which is the proposed side of VP2 contact in the DLP structure (55). The chimera PRE1+2 included the divergent EC2 site; both EC1 and EC3 were conserved between the polymerases and, thus, there was no rationale for altering them (**Fig. 3-8**). Unfortunately, none of the engineered rVP1 chimeras were active in the presence of PO-13 (C4) rVP2. If fact, three of the chimeras actually lost their capacity to be activated by the 'promiscuous' SA11 (C5) rVP2, indicating that the polymerases might adopt conformations that no longer are capable of interacting with VP2 in the manner necessary for activation (**Fig. 3-10**). This result was surprising to us because most of the residues we changed are surface-exposed and not implicated in supporting the intramolecular VP1 structure. For instance, the VL1 chimera swapped out a flexible helix-loop element on the surface of the VP1 NTD (**Figs. 3-8, 3-9, 3-10**). The PO-13 (R4) VL1 surface element is clearly functional in the context of a PO-13 (R4) backbone, but not on the SA11 (R2) backbone, at least in regard to activation by SA11 (C5) rVP2 (**Fig. 3-10**). The one chimera that did retain functionality with SA11 (C5) rVP2 (i.e., VL2 rVP1) also swapped out a non-flexible, surface-exposed helix-loop element (**Fig. 3-10**). However, this particular

swap (and the 12 amino acid changes wherein) was not sufficient to allow recognition by PO-13 (C4) rVP2. Ongoing experiments in our laboratories are assessing the role of VP1 +RNA exit tunnel interface residues by performing single and clustered point mutagenesis (see Chapter 4). Such work will further our understanding of whether or not the VP1-VP2 contact points observed in the DLP structure play important roles during dsRNA synthesis.

We also engineered chimeric rVP2 proteins that have swapped either the NTD or principal scaffold domains between SA11 (C5) rVP2 and PO-13 (C4) rVP2 (i.e., PO:SA and SA:PO). Consistent with a previous study using an RVA vs. RVC proteins, our results with PO:SA and SA11 (R2) rVP1 show that the specificity of rVP2 for a particular rVP1 lies in the principal scaffold domain, not the NTD (**Fig. 3-12**) (119). Yet, the result with SA:PO also shows that rVP2 proteins with a 'mismatched' NTD and principal scaffold domains are less robust in supporting polymerase activation than are those with matched domains (i.e., wildtype PO-13 (C4) rVP2). This result indicates that there are as of yet unknown intramolecular interactions between the NTD and the principal scaffold domain. It is possible that the presence and sequence of the NTD is important for proper folding of the principal scaffold domain, at least in regions that mediate VP1 binding. Because the VP1 contact sites on VP2 that are predicted in the DLP (EC; residues 377-403 of strain UK) are completely conserved among all genotypes, it is unlikely that they are involved in dictating polymerase specificity in the context of genome replication.

Altogether, the combined results of these bioinformatic analyses and *in vitro* assays raise questions about the possibility that VP1-VP2 contact points differ during transcription and replication and might even differ depending upon the precise genotype/strain combinations tested. Still, a limitation of our study is that we are using the *in vitro* dsRNA synthesis assay as

an indirect readout of functional rVP1-rVP2 interactions. At this time, we cannot exclude the possibility that enzymatic activation of rVP1 by rVP2 occurs via a mechanism that is independent of direct interactions between these proteins. For example, it is possible that rVP2 binds to the +RNA template, rather than to rVP1, and induces structural changes within the +RNA that allow it to bind to the polymerase in a manner that is 'in-register' with the catalytic site. In this case, polymerase activation would still depend upon the presence of the core shell protein but it would occur independent of direct rVP1-rVP2 protein interactions. Unfortunately, due to the natural propensity of the core shell protein to asymmetrically multimerize and the presence of co-purifying contaminants/proteolysis products, the rVP2 preparations are not amenable to standard protein-protein interaction assays (46, 198, 200). Future experiments in the lab will employ a virus like particle (VLP) approach, which relies on the co-expression of rVP1, rVP2, and rVP6 in insect cells and the formation of double-shelled capsids containing the polymerase, to detect rVP1-rVP2 interactions (46, 198, 201). We have previously used this approach to show that the RVC rVP1 (strain Bristol) failed to be efficiently packaged into VLPs made with RVA rVP2 (strain SA11) (201). For this protein pair set, the lack of VLP interaction correlated with a lack of *in vitro* activation in the dsRNA synthesis assay (119, 201). Indeed, future investigation into how the activity of the RVA polymerase is regulated by the core shell protein and whether or not VP1 activation requires VP2 binding are warranted and will enhance our understanding of a critical step in the viral replication cycle.

Materials and Methods

Intergenotypic bioinformatics analyses of VP1 and VP1. A total of 158 nucleotide sequences representing all available VP1 and VP2 genotypes (R1-R22 and C1-C20) were

obtained from NCBI (www.ncbi.nlm.nih.gov/nuccore) and trimmed to ORF boundaries (**Appendix A**). ORF nucleotide sequences were translated, aligned as amino acids using MAFFT with default settings, and untranslated using Python scripts (http://www.github.com/jbpease/mixtape) (202). At the time of the study, genotypes R17, R21, R22, C16, C19, and C20 had no available sequences in the database and were thus excluded from the study. However, most genotypes had 1-9 VP1 and VP2 ORF sequences available (**Appendix A**). For genotypes with >10 available full-length sequences (i.e., R1-R3, C1-C3), preliminary intragenotypic neighbor-joining phylogenetic trees were created using Geneious Pro v10.2.3 (BioMatters) to identify representative strains to use in intergenotypic analyses (**Appendix A**). Final, maximum-likelihood phylogenies were inferred from the aligned nucleotide ORF data using RAxML v.8.2 with the GTRGAMMA model and Rapid Bootstrapping (203). Monophyletic groupings were collapsed using FigTree v1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/).

Python scripts were used to calculate pairwise distances between sequences and multidimensional scaling plots were calculated using the scikit.sklearn module for Python (196). Python scripts were also used to determine amino acid alleles that were exclusively found in specific genotypic groups for each gene (scripts are available upon request). For visual representations of amino acid sequence alignments, Geneious Pro v10.2.3 (BioMatters) was used with the CLUSTALW plug-in and the BLOSUM cost matrix. Consensus sequences were derived using 0% majority threshold for translated intragenotypic amino acid alignments. Figure images were generated and colorized using Adobe Illustrator CS6 (Adobe Systems).

Cloning VP1 and VP2 genes and generation of recombinant baculoviruses. To create entry vectors for the generation of baculoviruses expressing DS-1 (R2) rVP1, Wa (R1)

rVP1, and ETD (R7) rVP1, cDNAs corresponding to the ORFs were codon-optimized for insect cell expression and the sequence for C-terminal 6XHis tags were added. The cDNAs were then synthesized *de novo* by GeneArt (Regensburg, Germany) and subcloned into pENTR-1A. The amino acid sequences of the DS-1 (R2) rVP1, Wa (R1) rVP1, and ETD (R7) rVP1 proteins exactly match sequences in the GenBank database (accession numbers DQ870505, DQ490539, and GQ479947, respectively). To create entry vectors for the generation of a baculovirus expressing PO-13 (R4) rVP1, outward PCR was performed to modify the pENTR-1A-PO-13 VP1 clone generated by McDonald and Patton (119), so that it would contain a Arg-Gly-Ser-Arg-Gly-Ser linker immediately preceding the 6XHis tag on the extreme C terminus. The addition of this linker improved the binding of PO-13 (R4) rVP1 to the metal affinity resin during purification. To create entry vectors for the generation of baculoviruses expressing DS-1 (C2) rVP2 and Wa (C1) rVP2, previously generated pCI vectors were used PCR templates for subcloning into the pENTR-1A (199). To create entry vectors for the generation of baculoviruses expressing the rVP1 chimeras (PRE1, PRE1+2, VL1, and VL2), blunt-end PCR was used. The vector sequence was amplified using nonphosphorylated primers and pENTR-SA11-VP1 as a template. The insert sequence was amplified using 5'-phosphorylated primers and pENTR-PO-13-VP1 as a template. In all PCR reactions Accuprime *Pfx* Supermix (Invitrogen) was used as the enzyme, PCR products were treated with DpnI (New England BioLabs), and cDNAs were gel purified prior to ligation using T4 DNA ligase (New England BioLabs). To create entry vectors for expression of rVP2 NTD-swap chimeras (SA:PO rVP2 and PO:SA rVP2), the cDNAs were codon-optimized for expression in insect cells, synthesized *de novo* by GeneArt (Regensburg, Germany), and subcloned into pENTR-1A. All final pENTR-

1A entry vector clones were all sequenced across the VP1/VP2 ORF prior to the generation of baculoviruses.

The BaculoDirect Expression System (Life Technologies) was used in accordance with the manufacturer's protocol to create recombinant baculoviruses expressing rVP1 or rVP2. Briefly, the VP1- and VP2-encoding genes in the pENTR-1A entry vectors were individually inserted into BaculoDirect C-Term linear DNA by recombination with LR Clonase II. The baculovirus DNA was then transfected into *Spodoptera frugiperda* (Sf9) cells using Cellfectin reagent (Life Technologies) and recombinant baculovirus was harvested from selective medium containing 100 μM ganciclovir. Sf9 cells were maintained at 28°C in complete Grace's medium (Life Technologies) supplemented to contain 10% fetal bovine serum, 100 U/mL penicillin, 100 μg/mL streptomycin, 0.5 μg/mL amphotericin B, and 1% Pluronic F-68 (Life Technologies). Baculoviruses expressing SA11 (R1) rVP1, SA11 (C5) rVP2, ETD (C7) rVP2, and PO-13 (C4) rVP2 were generated previously (119, 183).

Purification of rVP1 and rVP2 and *in vitro* dsRNA synthesis assays. His-tagged rVP1 and untagged rVP2 proteins were expressed as described previously (119) with some modifications. Briefly, Sf9 cells (2.5×10^7) were infected with the recombinant baculovirus and incubated in complete Grace's insect medium (Invitrogen) supplemented to contain 10% fetal bovine serum, 100 U/mL penicillin, 100 μg/mL streptomycin, 0.5 μg/mL amphotericin B, and 1% Pluronic F-68 (Life Technologies) while shaking at 115 rpm for 96 hours at 20°C or 25°C (VP1 or VP2, respectively). The rVP1 and rVP2 proteins were then purified from the Sf9 cells as described previously (119, 186). Purified rVP1 and rVP2 proteins were assessed for quality and relative quantity versus Precision Plus Protein Kaleidoscope Prestained Protein Standards

(Bio-Rad) in 4-15% SDS-PAGE gels (Protean) and Coomassie Blue staining (Thermo Scientific).

In vitro dsRNA synthesis assays were performed as described previously (118). Briefly, each 20-μL reaction mixture contained 2 pmol rVP1; 8 pmol rVP2; 16 pmol SA11 gene 8 +RNA; 50 mM Tris-HCl (pH 7.5); 1 μL 30% polyethyleneglycol 8000; 20 mM magnesium acetate; 1.6 mM manganese acetate; 2.5 mM dithiothreitol; 1.25 mM (each) ATP, CTP, and UTP; 5 mM GTP; 1 μL RNasin (Promega); and 1 μCi [α-^{32}P]UTP (3,000 Ci/mmol; PerkinElmer). Reactions proceeded at 37°C for 3 hours. The [^{32}P]-labeled dsRNA products of the reaction were electrophoresed in 4-15% SDS-PAGE gels (Protean) and visualized using a GE Healthcare Storm 860 phosphorimager or autoradiography. A minimum of 3 replicate experiments were performed from 2-4 batches of independently purified recombinant proteins (n=6-12 replicates total). Quantification of images obtained using the phosphorimager was completed using ImageQuant 5.2 software. For experiments including autoradiographs, ImageJ 1.49v was used to quantify band intensity of TIF images. One-sided *t* tests of the mean versus wildtype dsRNA levels set to hypothetical value of 100% were performed for quantified replicase assay data sets with Smith's Statistical Package, version 2.80. *P* values of <0.05 were considered statistically significant.

Molecular dynamics simulations of VP1 proteins. Molecular dynamics simulations were performed using GROMACS v5.1.3 on a modified atomic model of strain SA11-4F VP1 (PDB accession no. 2R7Q) or on a homology model of PO-13 VP1, which was created using the MODELLER interface to UCSF Chimera (43, 204, 205). Previously described modifications to the SA11-4F VP1 structure include a modeled flexible loop (residues 346-358) that is missing from the atomic structure (186). The PDB files of the modeled structures are available upon

request. Simulations were performed as described previously (206). Three trajectories initiated with different random seeds were run for both protein structures and the RMSF of α-carbons from each of the three trajectories was calculated using gmx rmsf command in GROMACS. B-factors (i.e., Debye-Waller factor) for each residue were calculated from the RMSF values using an established equation [B-factor=$8\pi^2/3 \times (RMSF^2]$. One-way ANOVA analyses were performed using StatPlus and *p* values <0.05 were considered statistically significant.

CHAPTER 4:

IN VITRO ACTIVITY OF THE ROTAVIRUS VP1 POLYMERASE IS MEDIATED BY SELECT VP2 CORE SHELL CONTACT SITES

Courtney L. Steger, Crystal E. Boudreaux, Courtney A. Cohen, and Sarah M. McDonald

This article is in preparation for submission to the Journal of Virology.

CLS and SMM wrote the manuscript and generated the figures. CEB and CAC generated preliminary data that inspired the development of this project.

Abstract

The rotavirus polymerase (VP1) mediates all stages of viral RNA synthesis within the confines of subviral particles and while associated with the core shell protein (VP2). Transcription [positive-strand (+) RNA synthesis] by VP1 occurs within double-layered particles (DLPs), while genome replication [double-stranded (ds) RNA synthesis] by VP1 occurs within assembly intermediates. VP2 is critical for the enzymatic activation of VP1; yet the mechanism by which the core shell protein triggers polymerase function remains poorly understood. Structural analyses of transcriptionally-competent DLPs show that VP1 is located beneath the VP2 core shell and sits slightly off-center from each of the icosahedral fivefold axes. In this position, the polymerase is contacted by VP2 at five distinct surface-exposed regions, comprising VP1 residues 264-267, 547-550, 614-620, 968-980, and 1022-1025. Here, we sought to test the

functional significance of these VP2 contact sites on VP1 activity using recombinant (r) proteins and an *in vitro* dsRNA synthesis assay. Specifically, we engineered 19 rVP1 proteins that contained single- or multi-point alanine mutations within each of the five sites. We assayed the mutant rVP1 proteins for the capacity to synthesize dsRNA *in vitro* in the presence of rVP2 and compared the results to those using wildtype controls. Our results identified one single-point mutant (R614A) and two multi-point mutants (E265A/L267A and D971A/S978A/I980A) that showed decreased rVP2-dependent polymerase activity. These results suggest that core shell protein binding to these three sites on the polymerase may be necessary for its activation during particle-associated RNA synthesis.

Importance

Rotaviruses are important pathogens that cause severe gastroenteritis in the young of many animals. The rotavirus VP1 polymerase mediates all stages of viral RNA synthesis, and it requires the VP2 core shell protein for its enzymatic function. Yet, there are several gaps in knowledge about how VP2 engages and activates VP1. Here, we probed the functional significance of five distinct VP2 contact sites on VP1 that were revealed through previous structural studies. Specifically, we engineered mutations within each of the five regions and assessed the mutant VP1 proteins for the capacity to synthesize RNA in the presence of VP2 in a test tube. Our results identified three surface-exposed VP1 regions that are critical for robust VP2-induced polymerase activity. These results are important because they enhance an understanding of a critical aspect of the rotavirus replication cycle and may inform disease treatment and prevention measures.

Introduction

Rotaviruses (RVs) are segmented, double-stranded (ds) RNA viruses that cause severe, life-threatening gastroenteritis in children and other animals (6, 176). The RV RNA-dependent RNA polymerase (VP1) catalyzes both the transcription of positive-strand (+) RNAs and the replication of dsRNA genome segments in a particle-associated manner (176). Specifically, VP1 is a structural protein, and 11-12 copies are tethered beneath the innermost icosahedral core shell of the triple-layered RV particle (**Fig. 4-1A-B**) (46, 47, 49, 51, 55). A single VP1 monomer is located just off-center from each of the core shell's icosahedral fivefold axes (**Fig. 4-1B**) (55). The core shell itself is comprised of 120 copies of VP2 that are arranged as asymmetric dimers, wherein one structural conformer (VP2-A) converges around the fivefold axes, and a different conformer (VP2-B) interdigitates between adjacent VP2-A proteins (**Fig. 4-1A-B**) (46, 47, 49, 51). At least three individual VP2 monomers of a fivefold decameric unit (i.e., VP2-A1, VP2-A2, and VP2-B1) make contacts with VP1 in this position (**Fig. 4-1B-C**) (55). It is largely thought, albeit not experimentally proven, that VP1 functions as a polymerase while tethered to the VP2 core shell at this site. However, the role of the core shell protein goes beyond merely serving as a structural scaffold for the polymerase. In particular, the capacity of recombinant (r) VP1 to synthesize dsRNAs *in vitro* is dependent upon the presence of rVP2 in reactions mixtures (42, 118, 119). The current hypothesis is that VP2 engages VP1, thereby triggering structural changes within the enzyme that allow it to function as a polymerase. However, the mechanistic details of VP1-VP2 interaction(s) that culminate in polymerase activation and particle-associated RNA synthesis are poorly understood.

VP1 mediates synthesis of +RNA transcripts as well as dsRNA genome segments while in conjunction with VP2 and within the confines of subviral particles (207). More specifically,

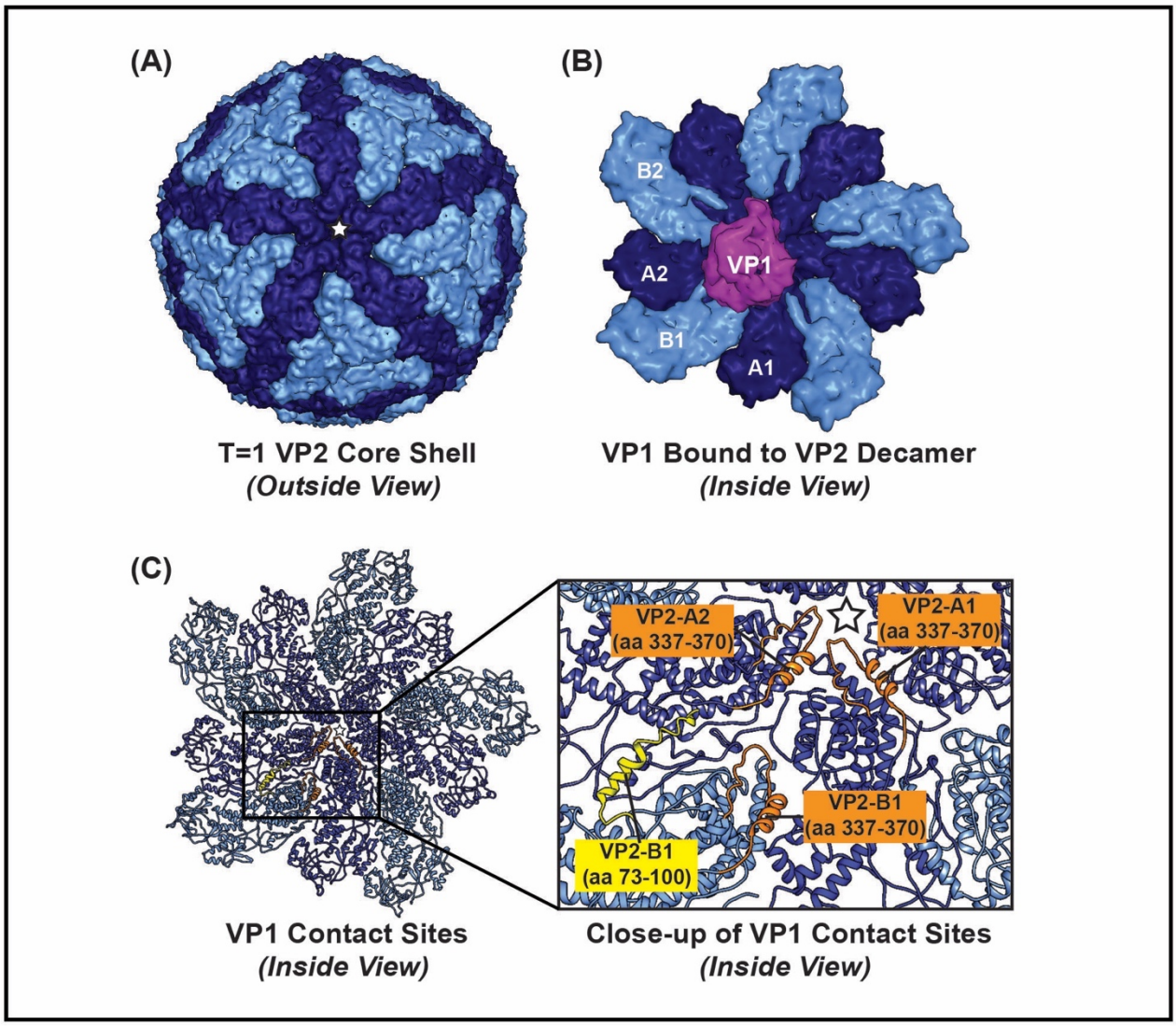


Fig. 4-1. Structure of the VP2 core shell and position of VP1. (A) Surface rendering of the VP2 core shell (PDB accession no. 4F5X), viewed from outside the particle into an icosahedral fivefold axis (star). Each fivefold axis is tightly encircled by five copies of VP2-A conformers (dark blue), while five VP2-B conformers (light blue) intercalate between VP2-A. Twelve decameric units of VP2 surround each fivefold axis and comprise the innermost core shell. (B) Internal view of a fivefold decameric unit, represented in surface fill with colors for VP2-A and VP2-B as in panel A. Multiple individual VP2 monomers (e.g., A1, A2, and B1) of the decamer contact a single VP1 monomer (magenta) positioned below the core shell just off-center from the fivefold axis, which is obscured by VP1 in this image. (C) Ribbon representation of the VP2 decamer shown in panel B is shown on the left. For clarity, the VP1 monomer has been removed. A black box outlines the area that is magnified in the image on the right. Regions of VP2 that contact VP1 in within the DLP are highlighted in orange (UK aa 337-370) or bright yellow (UK aa 73-100) for three distinct VP2 monomers (A1, A2, and B1). The fivefold axis is represented by a star, as in panel A.

during transcription, VP1 synthesizes +RNAs in the context of double-layered particle (DLPs), which are virions that have shed their outermost layer. The +RNA transcripts are extruded from the DLP via aqueous pores at the fivefold axes (46, 87). In contrast, genome replication (dsRNA synthesis) by VP1 occurs in tandem with the early stages of virion particle assembly. It is thought that +RNAs are packaged into VP1/VP2-containing subviral assembly intermediates where they are subsequently converted to dsRNAs via minus-strand RNA synthesis (87, 106, 116). Although relatively little is known about the structure and composition of these assembly intermediates, all available data suggests that multimeric units of VP2 could engage VP1 monomers in this context (116, 119, 120). Core shell-dependent dsRNA synthesis by the RV polymerase can be recapitulated *in vitro* using rVP1 and rVP2, providing a robust platform for mapping the enzymatic activation determinants within both proteins (42, 118-120).

The crystal structure of a simian RV VP1 (strain SA11) has been solved to atomic resolution, revealing a globular, cage-like enzyme 1089 amino acids (aa) in length. VP1 is comprised of an amino-terminal domain (NTD: aa 1-332), a central right-handed polymerase domain (PD: aa 333-778), and a carboxy-terminal domain (CTD: aa 779-1089) (**Fig. 4-2A**) (43). The PD is structurally and functionally conserved across diverse viral RNA-dependent RNA polymerases, and it contains several orthologous subdomains (fingers, palm, and thumb) and motif elements (motifs A-F) (43, 182). The active site is buried within the hollow center of the globular enzyme, and it is comprised of invariant 'Glu-Asp-Asp' residues of the PD (43, 208). Four tunnels permeate VP1 and allow for the entry of single-stranded RNA templates, the entry of NTPs and divalent metal cations, the exit of +RNA transcripts, and the exit of newly formed dsRNA genome segments (**Fig. 4-2B-C**) (43). In the structure of the DLP (i.e., transcriptase complex), VP1 is oriented such that its +RNA exit tunnel borders the VP2 core shell. This

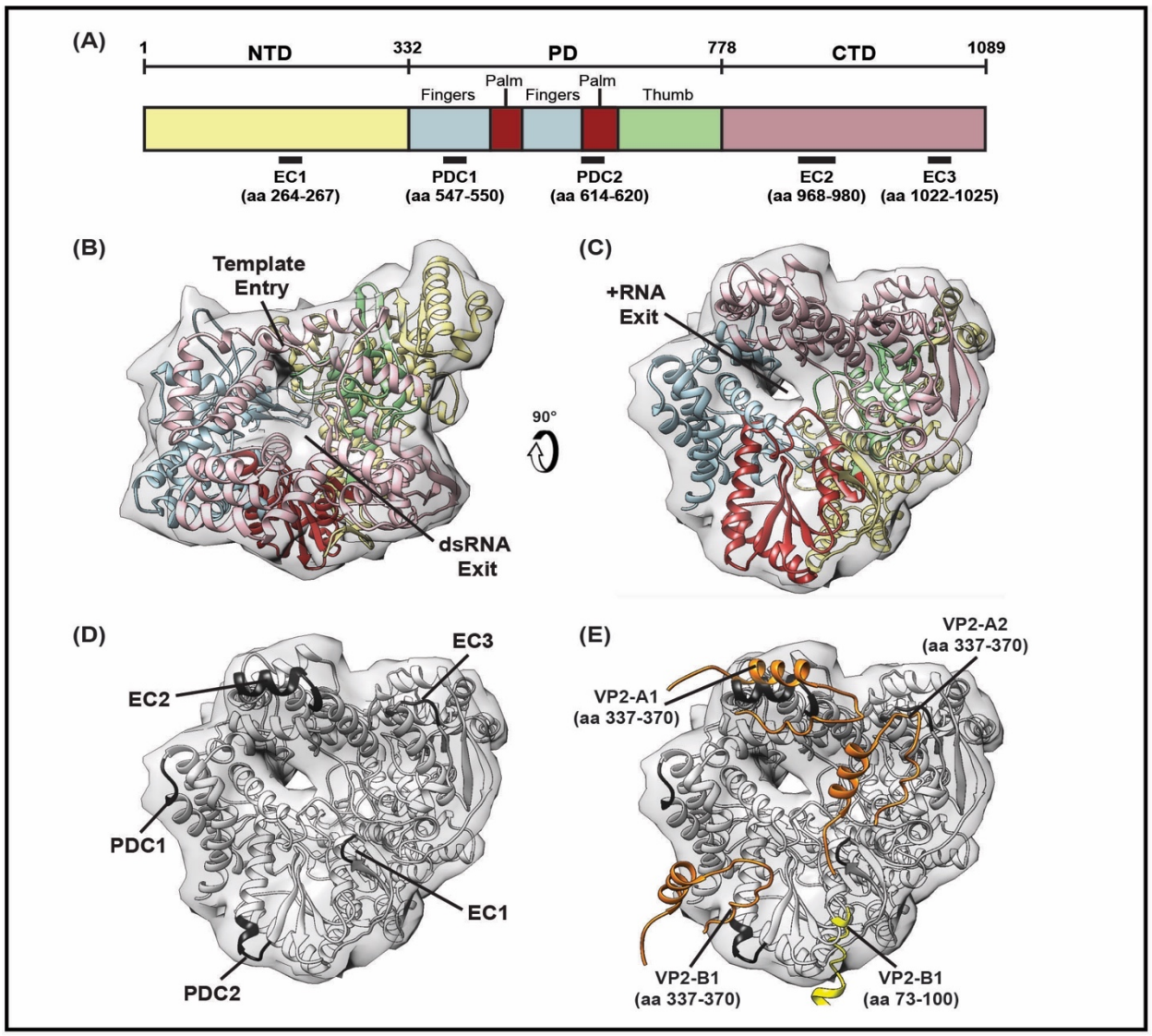


Fig. 4-2. Structure of VP1 and sites of contact with VP2. (A) VP1 gene shown as a linear schematic and colored according to domain/subdomain organization, with amino acid residues of domain interfaces labeled above the gene. The N-terminal domain (NTD; pale yellow) and C-terminal domain (CTD; pink) flank the central polymerase domain (PD), which is comprised of canonical finger (blue), palm (red), and thumb (green) subdomains. Regions of VP1 that contact VP2 in the DLP are labeled with a black bar below the corresponding region of the gene. (B) Conventional orientation of the VP1 enzyme (PDB accession no. 4AU6 and 4F5X) shown in ribbon representation and colored according to domain organization depicted in panel A, with transparent surface fill outlined in gray. Known or putative functions of two VP1 tunnels that traverse into the catalytic center are labeled. (C) Image in panel B rotated 90° backward to show the surface of VP1 surrounding the +RNA exit tunnel, which borders VP2 within the DLP. (D) Structure in panel C shown in grayscale, with known or putative VP1-VP2 contact sites on VP1 colored in black and listed on the image. Putative contact regions are labeled on the image as Estrozi contacts (EC) or polymerase domain contacts (PDC). (E) Structure in panel D with known or putative VP1-VP2 contact sites on three VP2 monomers (A1, A2, and B1) shown in ribbon representation and colored as in Fig. 4-1C. For clarity, all other regions of VP2 have been removed.

position and orientation for the RV polymerase is consistent with that of other *Reoviridae* family members (orthoreovirus, cypovirus, and aquareovirus) (37, 38, 117). For RV, Estrozi and colleagues identified three regions of VP1 that clash with (i.e., overlap) VP2 in the DLP structure (55). For clarity, we will refer to these regions as Estrozi Contact (EC) sites 1, 2, and 3. EC1 is located within the VP1 NTD (SA11 aa 264-267), while EC2 and EC3 are located in the VP1 CTD (SA11 aa 968-980 and 1022-1025, respectively) (**Fig. 4-2A and D**). Two additional regions within the VP1 PD are surface-exposed and proximal to the core shell; we refer to these as PD Contact sites 1 and 2 (PDC1 and PDC2). PDC1 is positioned within the fingers subdomain of the PD (SA11 aa 547-550), whereas PDC2 is situated in the palm subdomain of the PD (SA11 aa 614-620) (**Fig. 4-2A and D**). On the VP2 side of the interaction, two VP2-A subunits (A1 and A2) and one VP2-B subunit (B1) make contacts with VP1 (**Fig. 4-1C and Fig. 4-2E**) (55). The most extensive overlap occurs within a region of VP2 that encompasses aa 337-370 (bovine strain UK) within the principal scaffold domain of the core shell. Moreover, the extreme amino-terminal domain of VP2 (aa 1-100) is predicted to protrude downward into the core and contact VP1, but this region is not fully-resolved in the existing structures. Still, residues 73-100 of a single VP2-B conformer (B1) converge towards VP1 EC1 in the DLP structure (**Fig. 4-1C** and **Fig. 4-2E**).

In the current study, we probed the functional importance of the five VP2 contact sites on VP1 that were identified from the DLP structure (i.e., EC1, PDC1, PDC2, EC2, and EC3) using alanine mutagenesis and the *in vitro* dsRNA synthesis assay. Our results show that alanine mutations at some residues in VP1 regions EC1, PDC2, and EC2 reduce the capacity of rVP1 to synthesize dsRNA in the presence of rVP2. These results support the notion that core shell binding to these sites may be critical for VP1 enzymatic activation. Based on these results, we

speculate that the manner in which VP2 binds to VP1 during transcription (i.e., in the DLP) may be identical to that during genome replication (i.e., in assembly intermediates).

Results

EC1 residues mediate *in vitro* dsRNA synthesis. EC1 (SA11 VP1 aa 264-267: NEEL) is located in the VP1 NTD, and it contains two consecutive negatively-charged glutamic acid residues, flanked on either side by a polar residue (**Fig. 4-3A**). In general, these amino acids are conserved among group A RV strains with the exception of asparagine at position 264, which is threonine in some avian strains (**Fig. 4-3A**). The glutamic acid side chain at position 265 is both surface-exposed and points outwards towards the VP2 core shell, making contacts with aa 337-339 of a single VP2-A2 monomer (**Fig. 4-3A**). The structurally-resolved portion of the amino-terminal domain of a neighboring VP2-B1 monomer (aa 73-100) also protrudes towards EC1 and is predicted to clash with the polymerase at/near this site (**Fig. 4-3A**) (55). To investigate the significance of residues within EC1 on VP2-dependent VP1 polymerase activity, we engineered single and double alanine mutations at positions 265 and/or 267 (i.e., E265A and E265A/L267A) in the backbone of SA11 rVP1. The His-tagged E265A and E265A/L267A rVP1 and wildtype (WT) control rVP1 proteins were expressed in insect Sf9 cells using baculovirus, purified by metal-affinity chromatography, and visualized following SDS-PAGE and Coomassie Blue staining (**Fig. 4-3B**). The rVP1 proteins were then assayed for their capacity to synthesize dsRNA *in vitro* in reactions containing SA11 rVP2, SA11 +RNA template, divalent cations, NTPs, and trace amounts of ^{32}P-UTP. The newly synthesized ^{32}P-labeled dsRNA products of the reactions were resolved by SDS-PAGE and then visualized/quantified using a phosphorimager (**Fig. 4-3C-D**). The results showed that E265A rVP1 synthesized dsRNA at levels that were

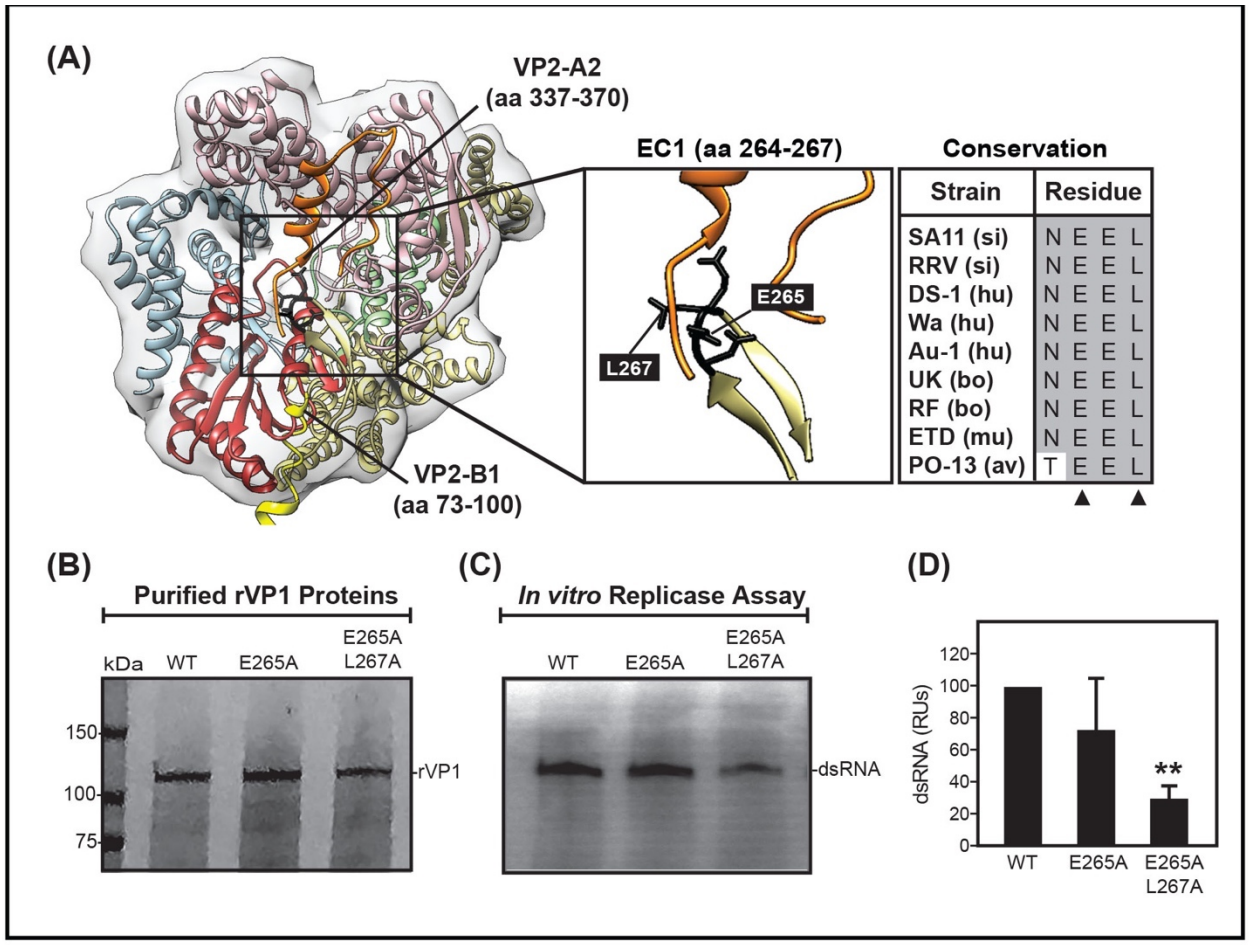


Fig. 4-3. Characterization of VP1 region EC1. (A) A ribbon drawing of VP1 (PDB accession no. 4AU6 and 4F5X) is shown on the left in the same orientation and coloration as in Fig. 4-2C. Ribbon drawings of VP2-A2 (UK VP2 aa 337-370) and VP2-B1 (UK VP2 aa 73-100) are shown in the image on the left in the same orientation and coloration as in Fig. 4-2E. A black box outlines the area that is magnified in the image on the right. For clarity, some regions of VP1 and VP2 have been removed. Side chains are shown in black for region EC1 (SA11 VP1 aa 264-267: NEEL) and residues mutated in this study are labeled. Amino acid sequence conservation for EC1 is shown as an alignment in which strain name is listed to the left, with species of origin in parentheses; gray shading indicates conservation of amino acid identity and variable amino acids are highlighted in white. (B) Approximately 2 pmol purified wildtype SA11 rVP1 or point mutants E265A rVP1, and E265A/L267A rVP1 were electrophoresed in a 4-15% SDS-polyacrylamide gel and visualized by Coomassie Blue stain. A molecular mass (in kilodaltons) marker is shown to the left. (C) Radiolabeled dsRNA synthesis products synthesized by 2 pmol of each rVP1 in the presence of 8 pmol of rVP2 from strain SA11 rVP2. All reactions contained 8 pmol of an RVA +RNA template and were incubated at 37°C for 180 minutes. Radiolabeled dsRNA products were resolved in 4-15% SDS polyacrylamide gels and visualized using a phosphorimager. (D) Radiolabeled dsRNA from 6 independent experiments using at least 3 protein batches was quantified and expressed as relative units (RUs). Averages are shown as bar graphs below each gel, and error bars represent standard deviations from the mean. A double asterisk indicates a P value <0.001.

indistinguishable from WT rVP1. However, the E265A/L267A rVP1 double-alanine mutant synthesized significantly-reduced levels of dsRNA product, yielding only ~25% of WT levels (**Fig. 4-3D**). These results indicate that mutating the highly conserved, surface-exposed, electrically-charged glutamic acid amino acid residue to a hydrophobic alanine residue alone was not sufficient to abrogate rVP2 interactions and/or diminish enzymatic activity of rVP1. However, combining this E265A mutation with another more modest leucine-to-alanine changed at position 267 (i.e., L267A) was sufficient to decrease *in vitro* enzymatic activity of rVP1 in the presence of rVP2.

PDC1 residues are not critical for *in vitro* dsRNA synthesis. PDC1 (SA11 VP1 aa 547-550: NMTN) is located in the fingers subdomain of the PD, and it maintains highly conserved amino acid chemistry at each position (**Fig. 4-4A**). In particular, this region contains three polar residues (N547, T549, and N550), and one hydrophobic residue (M548) in the middle (**Fig. 4-4A**). The asparagine at position 547 is conserved among most group A RV strains, but can be threonine in some avian and/or murine strains (**Fig. 4-4A**). Also, some murine RV strains show a unique methionine-to-leucine mutation at position 548 (**Fig. 4-4A**). Residues M548 and N550 have side chains that are pointed outwards towards the VP2 core shell, specifically towards aa 373-370 of VP2-B1, though they do not directly clash with any of the VP2 monomers in the DLP structure (**Fig. 4-4A**). To investigate whether these PDC1 residues were critical for VP2-dependent VP1 polymerase activity, we engineered and purified rVP1 mutants with individual alanine mutations at each position (i.e., N547A, M548A, T549A, and N550A) (**Fig. 4-4B**). Our rationale for making only single-alanine mutations was that this region lies in a subdomain that is important for a catalysis; thus, multiple mutations in this region would be expected to negatively impact the enzymatic activity of the polymerase in a manner that is independent of VP2 contacts.

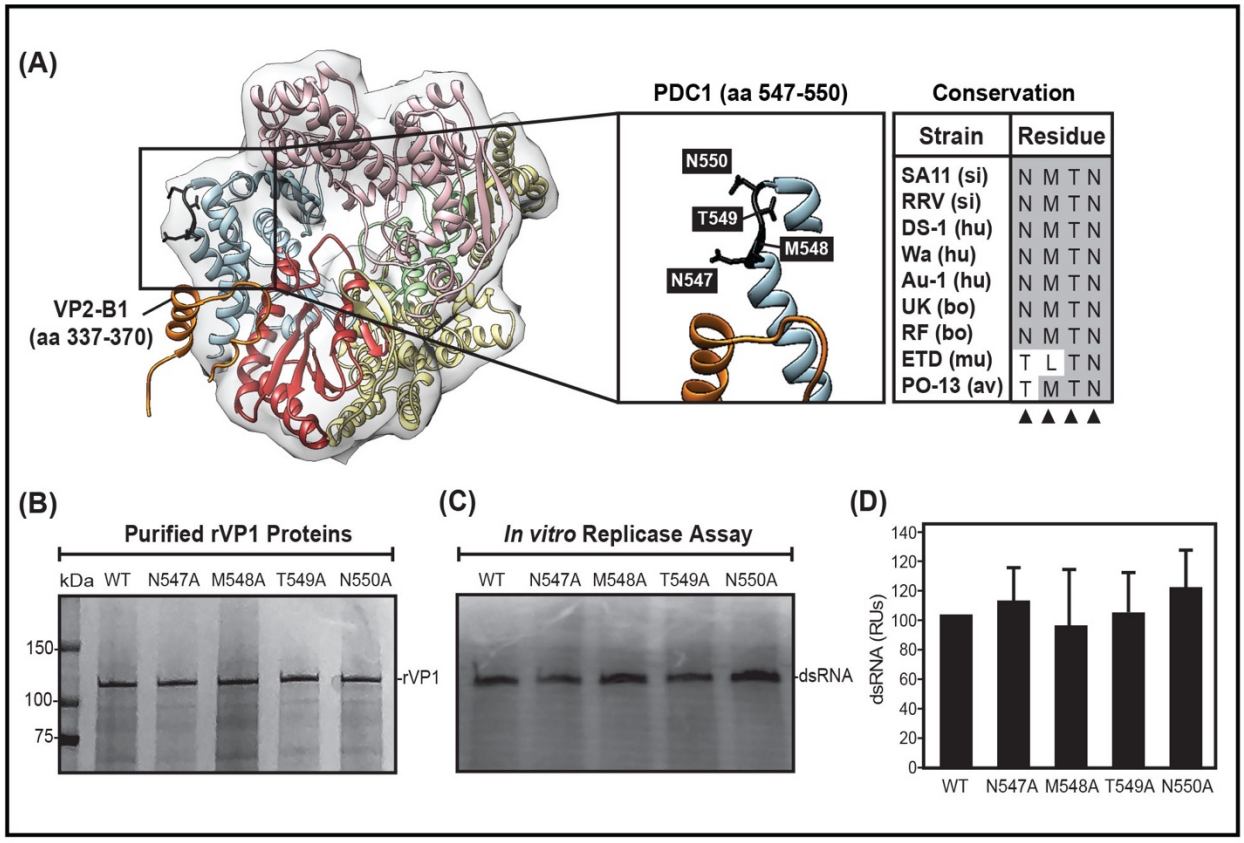


Fig. 4-4. Characterization of VP1 region PDC1. (A) A ribbon drawing of VP1 (PDB accession no. 4AU6 and 4F5X) is shown on the left in the same orientation and coloration as in Fig. 4-2C. A ribbon drawing of VP2-B1 (UK VP2 aa 337-370) is shown in the image on the left in the same orientation and coloration as in Fig. 4-2E. A black box outlines the area that is magnified in the image on the right. For clarity, some regions of VP1 and VP2 have been removed. Side chains are shown in black for region PDC1 (SA11 VP1 aa 547-550: NMTN) and residues mutated in this study are labeled. Amino acid sequence conservation for PDC1 is shown as an alignment in which strain name is listed to the left, with species of origin in parentheses; gray shading indicates conservation of amino acid identity and variable amino acids are highlighted in white. (B) Approximately 2 pmol purified wildtype SA11 rVP1 or single point mutants N547A rVP1, M548A rVP1, T549A rVP1, and N550A rVP1 were electrophoresed in a 4-15% SDS-polyacrylamide gel and visualized by Coomassie Blue stain. A molecular mass (in kilodaltons) marker is shown to the left. (C) Radiolabeled dsRNA synthesis products synthesized by 2 pmol of each rVP1 in the presence of 8 pmol of rVP2 from strain SA11 rVP2. All reactions contained 8 pmol of an RVA +RNA template and were incubated at 37°C for 180 minutes. Radiolabeled dsRNA products were resolved in 4-15% SDS-polyacrylamide gels and visualized using a phosphorimager. (D) Radiolabeled dsRNA from 4 independent experiments using at least 2 protein batches was quantified and expressed as relative units (RUs). Averages are shown as bar graphs below each gel, and error bars represent standard deviations from the mean.

The results showed that the *in vitro* dsRNA levels produced by PDC1 rVP1 mutants were comparable to WT rVP1 (**Fig. 4-4C-D**), suggesting that these rVP1 mutants were properly folded, interacted with rVP2, and were fully functional under the assay conditions. These data suggest that the uncharged polar or hydrophobic amino acid side chains within PDC1 individually are not critical for VP2 interaction(s) and enzyme function.

A single PDC2 residue mediates *in vitro* dsRNA synthesis. PDC2 (SA11 VP1 aa 614-620: RISNKHS) is located within the palm subdomain of the PD, and it contains two electrically-charged residues (i.e., R614 and K618), two hydrophobic residues (i.e., I615 and Y619), one highly conserved polar residue (i.e., N617) and several positions that accommodate variable amino acid chemical propensities (**Fig. 4-5A**). Most strikingly, position 618 contains a positively-charged lysine in most strains, but adopts a negatively-charged glutamic acid residue in a murine ETD RV strain (**Fig. 4-5A**). Additionally, the highly conserved hydrophobic tyrosine residue at position 619 is a positively-charged histidine residue in simian RV strain SA11, thereby altering the chemical properties of the residue, but maintaining the bulky ring structure in the amino acid side chain (**Fig. 4-5A**). Amino acids at positions 616 and 620 are not conserved, but rather they toggle between polar, uncharged serine residues or hydrophobic alanine amino acid residues for various RV strains (**Fig. 4-5A**). PDC2 contains some amino acids with electrically-charged side chains that are directed outwards towards the VP2 core shell, such as R614 and K618 (**Fig. 4-5A**). PDC2 residues do not make direct contact with the VP2 core shell in the structure, but are located adjacent to the loop-helix-loop regions of VP2-B1 (aa 373-370) in the DLP (**Fig. 4-5A**). To investigate whether these PDC2 residues were important for VP2-dependent VP1 polymerase activity, we engineered SA11 rVP1 mutants with single- or double-alanine mutations at positions that were highly conserved and/or electrically-charged

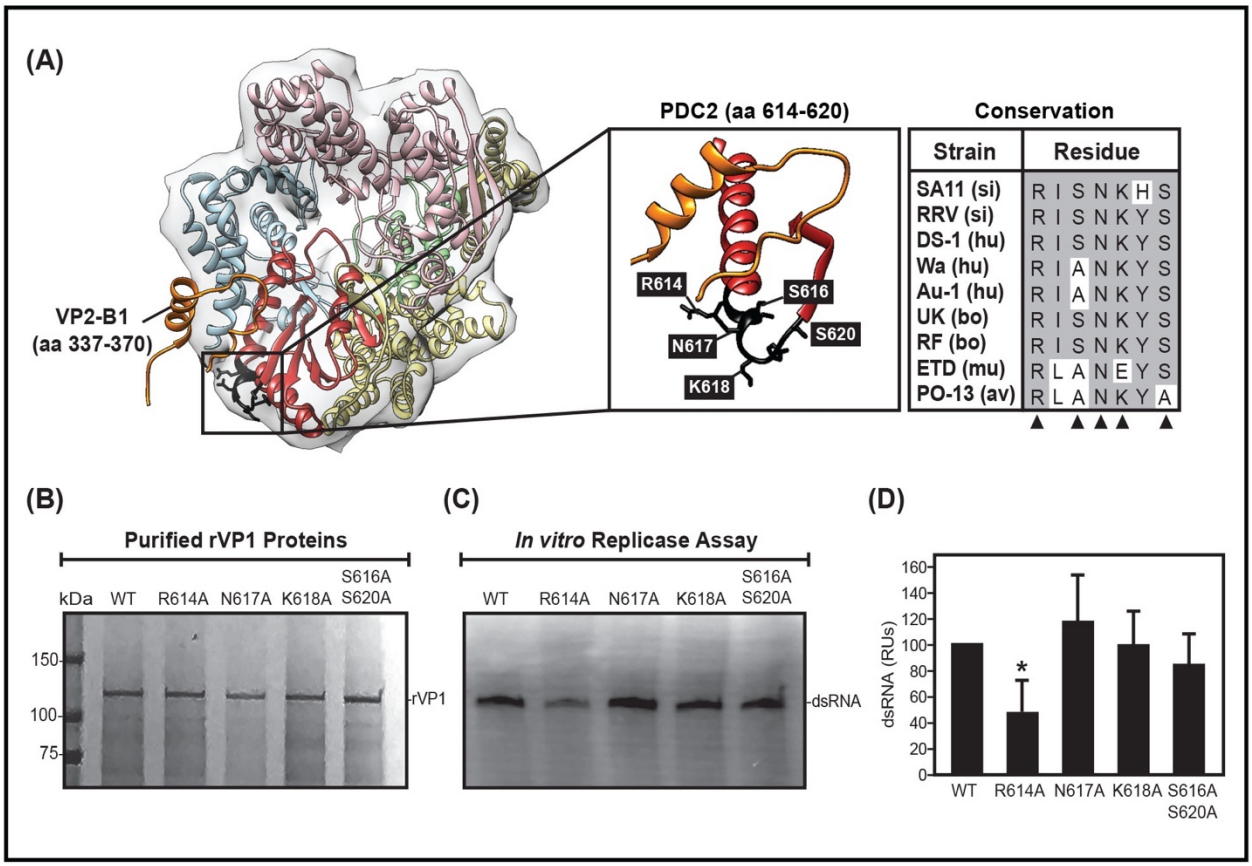


Fig. 4-5. Characterization of VP1 region PDC2. (A) A ribbon drawing of VP1 (PDB accession no. 4AU6 and 4F5X) is shown on the left in the same orientation and coloration as in Fig. 4-2C. A ribbon drawing of VP2-B1 (UK VP2 aa 337-370) is shown in the image on the left in the same orientation and coloration as in Fig. 4-2E. A black box outlines the area that is magnified in the image on the right. For clarity, some regions of VP1 and VP2 have been removed. Side chains are shown in black for region PDC2 (SA11 VP1 aa 614-620: RISNKHS) and residues mutated in this study are labeled. Amino acid sequence conservation for PDC2 is shown as an alignment in which strain name is listed to the left, with species of origin in parentheses; gray shading indicates conservation of amino acid identity and variable amino acids are highlighted in white. (B) Approximately 2 pmol purified wildtype SA11 rVP1 or point mutants R614A rVP1, N617A rVP1, K618A rVP1 and S616A/S620A rVP1 were electrophoresed in a 4-15% SDS-polyacrylamide gel and visualized by Coomassie Blue stain. A molecular mass (in kilodaltons) marker is shown to the left. (C) Radiolabeled dsRNA synthesis products synthesized by 2 pmol of each rVP1 in the presence of 8 pmol of rVP2 from strain SA11 rVP2. All reactions contained 8 pmol of an RVA +RNA template and were incubated at 37°C for 180 minutes. Radiolabeled dsRNA products were resolved in 4-15% SDS-polyacrylamide gels and visualized using a phosphorimager. (D) Radiolabeled dsRNA from 4 independent experiments using at least 2 protein batches was quantified and expressed as relative units (RUs). Averages are shown as bar graphs below each gel, and error bars represent standard deviations from the mean. A single asterisk indicates a P value <0.005.

(i.e., R614A, N617A, K618A, and S616A/S620A). Since PDC2 is located within the palm subdomain and neighbors the buried active site (SA11 aa residues 630-632), we reasoned that multiple mutations in this region could inhibit enzymatic activity in a VP2-independent manner (e.g., local conformational changes that disrupt the active site) and thus made single point mutants, with the exception of S616A/S620A rVP1. The rVP1 point mutants were expressed and purified alongside WT rVP1 control protein and assayed for in vitro dsRNA synthesis (**Fig. 4-5B-C**). The results showed that dsRNA levels produced by the mutants were comparable to that of WT, with the exception of R614A rVP1; this single alanine mutant produced dsRNA at levels that were only ~50% of WT rVP1 (**Fig. 4-5C-D**). These results suggest that mutating a highly conserved, surface-exposed, electrically-charged arginine amino acid residue to a hydrophobic alanine residue at position 614 was sufficient to diminish enzymatic activity of rVP1. Since the arginine side chain is surface-exposed and outward-facing, it is likely that rVP1 activity may be inhibited due to disrupted intermolecular interactions with rVP2. In contrast, dsRNA product levels from K618A rVP1, a position that is not completely conserved, were not significantly different from WT dsRNA levels (**Fig. 4-5C-D**). These results suggest that amino acid conservation within regions that are both surface-exposed and located on the VP1-VP2 interface is indicative of residues that are critical for maintaining enzyme function. Given that N617A rVP1, wherein a conserved, polar, uncharged asparagine side chain was mutated to alanine, or S616A/S620A rVP1, with hydrophobic serine residues changed to alanine, showed no significant loss of function phenotype, we conclude that mutating uncharged, albeit highly conserved, amino acids is not sufficient to disrupt rVP1 function during dsRNA synthesis.

Multiple EC2 residues are required for *in vitro* dsRNA synthesis. EC2 (SA11 VP1 aa 968-980: PKIDADTYVGSKI) is located in the VP1 CTD, and it directly clashes with aa 340-

370 of VP2-A1 in the structure of the DLP (**Fig. 4-6A**). EC2 is the largest of the VP2 contact regions and contains four electrically-charged residues (i.e., K969, D971, D973, and K979), two polar residues (i.e., T974 and S978) and many hydrophobic and/or nonpolar residues (**Fig. 4-6A**). Interestingly, several positions within EC2 show amino acid signatures unique to avian strains, including two residues at positions 969 and 979, which exhibit negatively-charged residues in avian strains, but maintain either positively-charged or nonpolar residues in other RV strains (**Fig. 4-6A**). Several amino acids within EC2 have highly conserved identity across all RV strains (i.e., P968, D971, Y975, S978, and I980) and even more display conserved amino acid similarity (**Fig. 4-6A**). Yet only one residue, at SA11 VP1 position D971, is both conserved and electrically-charged (**Fig. 4-6A**). Unfortunately, SA11 rVP1 proteins with a single point mutation substituting D971A were expressed at very low levels in insect cells and could not be purified at levels high enough to test for functional activity in the dsRNA synthesis assay (data not shown). Nonetheless, we generated SA11 rVP1 mutants with single alanine point mutations at conserved positions P968, I970, and S978 and assayed for the capacity to synthesize dsRNA *in vitro* (**Fig. 4-6B-C**). The results showed that the mutants produced dsRNA at levels comparable to WT rVP1, suggesting that they were all properly folded, interacted with rVP2, and were functional under the assay conditions (**Fig. 4-6C-D**). These data indicate that, individually, the amino acid side chains of conserved residues within this region are not critical for VP2 interactions or VP1 activity. Alternatively, it is possible that multiple amino acids within this region engage VP2 in a combinatorial mechanism to maintain enzyme function.

We next generated SA11 rVP1 multi-point mutants wherein 2-4 clustered, conserved, and surface-exposed amino acid residues within EC2 were switched to alanine (**Fig. 4-6B**). We hypothesized that clustered alanine substitutions would diminish functional capacity of SA11

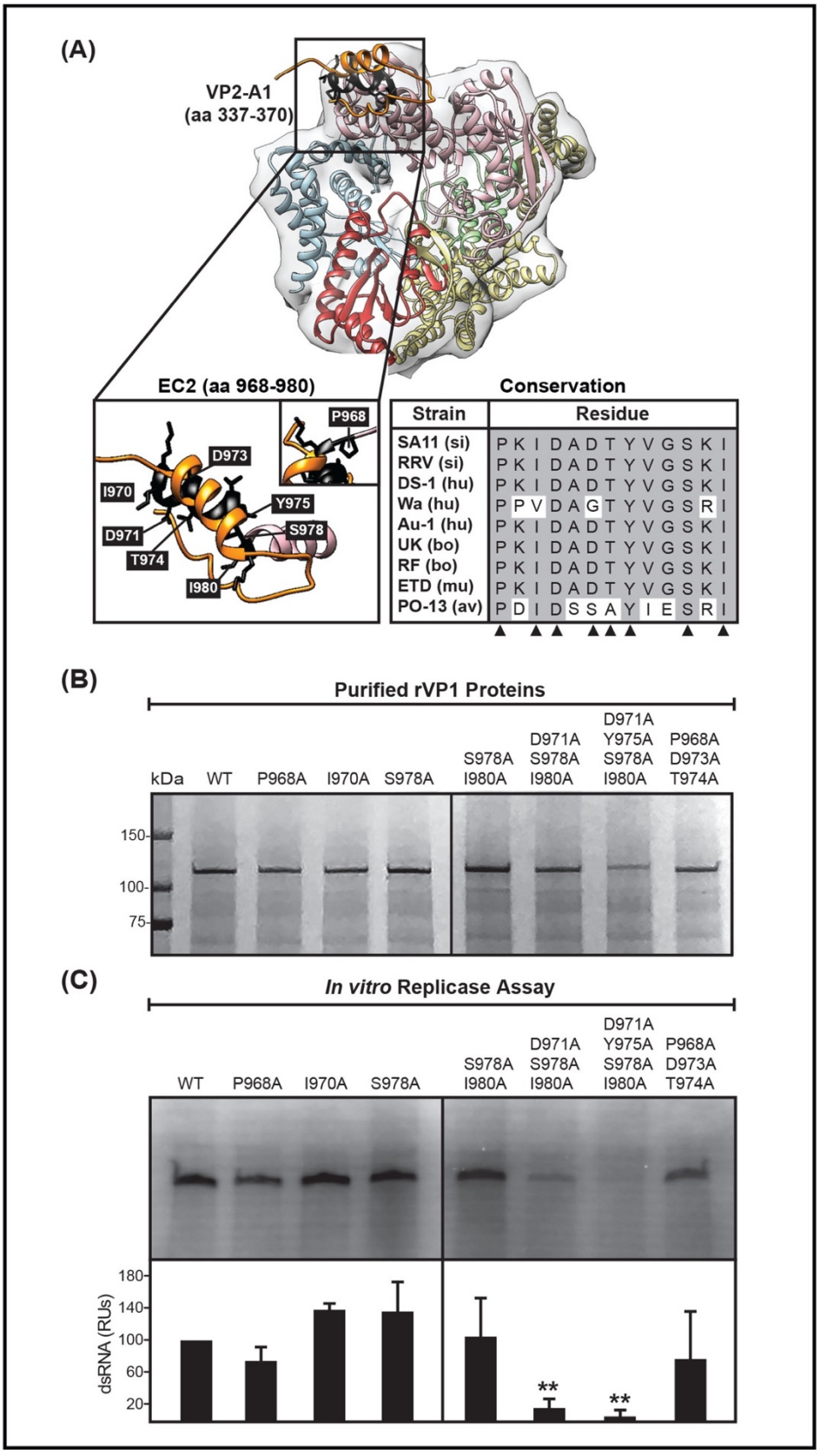


Fig. 4-6. Characterization of VP1 region EC2. Legend on following page.

Fig. 4-6. Characterization of VP1 region EC2. (A) A ribbon drawing of VP1 (PDB accession no. 4AU6 and 4F5X) is shown on the top in the same orientation and coloration as in Fig. 4-2C. A ribbon drawing of VP2-A1 (UK VP2 aa 337-370) is shown in the image on the bottom left in the same orientation and coloration as in Fig. 4-2E. A black box outlines the area that is magnified in the image on the top. For clarity, some regions of VP1 and VP2 have been removed and P968 is shown in the inset. Side chains are shown in black for region EC2 (SA11 VP1 aa 968-980: PKI-DADTYVGSKI) and residues mutated in this study are labeled. Amino acid sequence conservation for EC2 is shown as an alignment in which strain name is listed to the left, with species of origin in parentheses; gray shading indicates conservation of amino acid identity and variable amino acids are highlighted in white. (B) Approximately 2 pmol purified wildtype SA11 rVP1 or rVP1 mutants were electrophoresed in a 4-15% SDS-polyacrylamide gel and visualized by Coomassie Blue stain. A molecular mass (in kilodaltons) marker is shown to the left. (C) Radiolabeled dsRNA synthesis products synthesized by 2 pmol of each rVP1 in the presence of 8 pmol of rVP2 from strain SA11 rVP2. All reactions contained 8 pmol of an RVA +RNA template and were incubated at 37°C for 180 minutes. Radiolabeled dsRNA products were resolved in 4-15% SDS-polyacrylamide gels and visualized using a phosphorimager. Radiolabeled dsRNA from 6 independent experiments using at least 3 protein batches was quantified and expressed as relative units (RUs). Averages are shown as bar graphs below each gel, and error bars represent standard deviations from the mean. A double asterisk indicates a P value of <0.001.

rVP1 mutants if the changed amino acids were involved in maintaining enzymatic function via interactions with rVP2. Most of the multi-point mutant rVP1 proteins purified at levels comparable to SA11 rVP1 WT, with the exception of the quadruple point mutant (i.e., D971A/Y975A/S978A/I980A), which yielded lower levels of purified protein, possibly due misfolding (**Fig. 4-6B** and data not shown). Results of *in vitro* functional assays showed that mutants S978A/I980A rVP1 and P968A/D973A/T974A rVP1 produced dsRNA at levels comparable to WT rVP1 (**Fig. 4-6C-D**). In contrast, mutant D971A/S978A/I980A rVP1 showed significantly-reduced dsRNA levels, compared to WT rVP1, yielding only ~20% of the product (**Fig. 4-6C-D**). These results suggest that the addition of a D971A mutation in combination with S978A and I980A is sufficient to decrease dsRNA synthesis by rVP1 under the *in vitro* assay conditions. These results indicate that electrically-charged amino acid side chains of EC2, with or without combinatorial effects from surrounding amino acids, may be critical for maintaining functional capacity of rVP1 via interactions with rVP2 and/or for supporting enzyme function.

EC3 residues are not necessary for facilitating *in vitro* dsRNA synthesis. EC3 (SA11 VP1 aa 1022-1025: FKGK) is located in the VP1 CTD, and all four of its amino acid residues are completely conserved across group A RV strains (**Fig. 4-7A**). EC3 contains two positively-charged lysine residues with long amino acid side chains that extend out toward the VP2 core shell and overlap with aa 354-370 of VP2-A1 (**Fig. 4-7A**). We hypothesized that a multi-point mutant that contained a lysine-to-alanine substitution in combination with other alanine substitutions of surrounding amino acids would disrupt intermolecular interactions between rVP1 and rVP2, and thus decrease dsRNA synthesis levels. To test whether double or triple alanine point mutations were sufficient to diminish functional activity of the enzyme, we generated mutants with 2 or 3 alanine substitutions within EC3 (**Fig 4-7B**). Surprisingly, all of the rVP1 mutants produced dsRNA at levels indistinguishable from WT rVP1 (**Fig. 4-7C-D**). These results suggest that EC3 is not vital for rVP2 interactions or *in vitro* dsRNA synthesis by rVP1.

Discussion

RV RNA synthesis is mediated by the VP1 polymerase and occurs within the confines of subviral particles. The VP2 core shell protein serves as an internal scaffold for the polymerase and critically regulates VP1 function (42, 105, 118-120, 185, 186, 208). Specifically, it is hypothesized that VP2 directly engages surface-exposed residues of VP1, thereby inducing conformational changes that activate the polymerase to initiate RNA synthesis. This core shell-dependent polymerase regulatory mechanism serves to coordinate RNA synthesis with virion disassembly/assembly. However, there are several gaps in knowledge about how VP2 binds to and activates VP1. High-resolution structures of transcriptionally-competent DLPs have identified several surface-exposed sites on the VP1 +RNA exit tunnel interface that are contacted

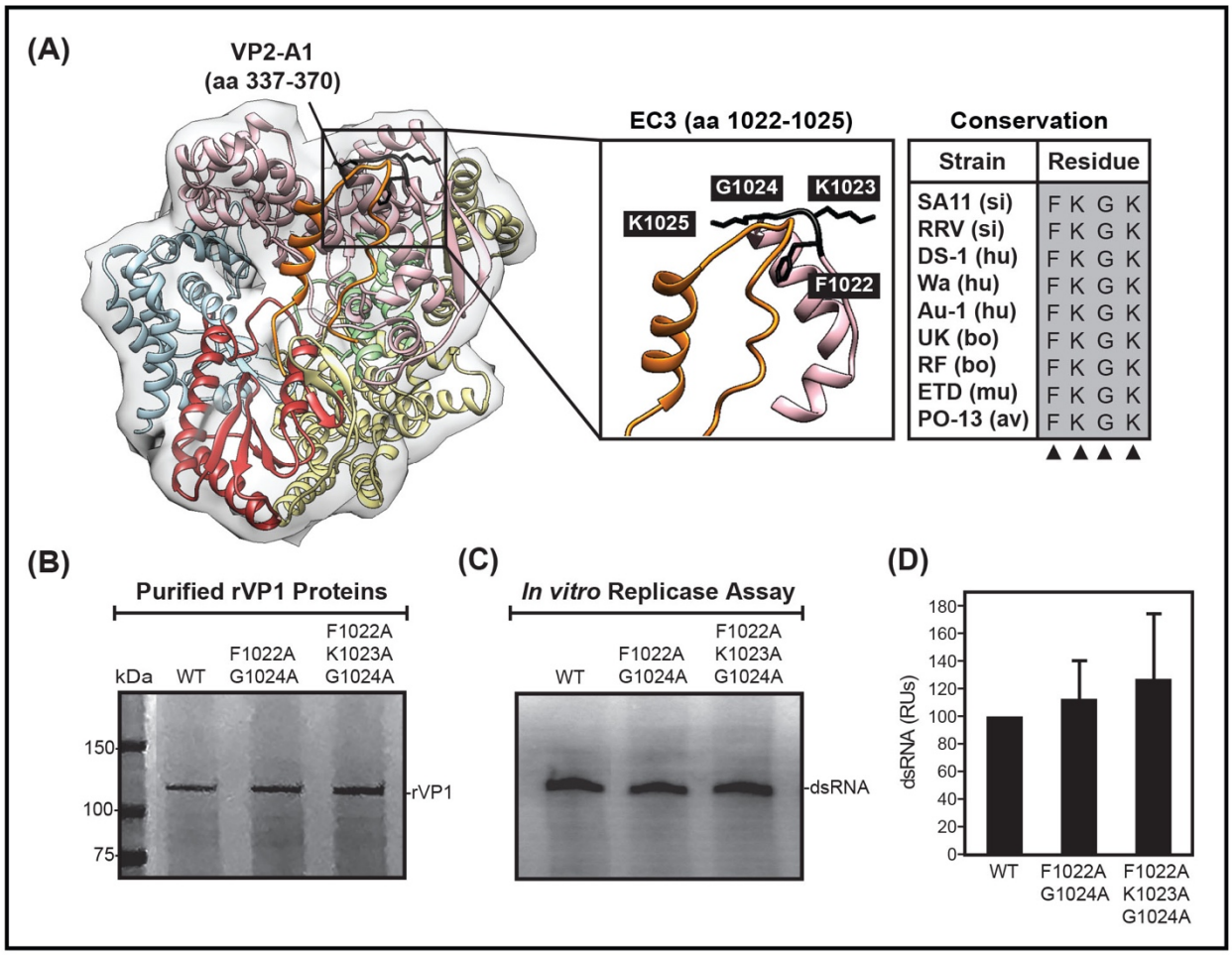


Fig. 4-7. Characterization of VP1 region EC3. (A) A ribbon drawing of VP1 (PDB accession no. 4AU6 and 4F5X) is shown on the left in the same orientation and coloration as in Fig. 4-2C. A ribbon drawing of VP2-A1 (UK VP2 aa 337-370) is shown in the image on the left in the same orientation and coloration as in Fig. 4-2E. A black box outlines the area that is magnified in the image on the right. For clarity, some regions of VP1 and VP2 have been removed. Side chains are shown in black for region EC3 (SA11 VP1 aa 1022-1025: FKGK) and residues mutated in this study are labeled. Amino acid sequence conservation for PDC1 is shown as an alignment in which strain name is listed to the left, with species of origin in parentheses; gray shading indicates conservation of amino acid identity. (B) Approximately 2 pmol purified wildtype SA11 rVP1 or multi-point mutants F1022A/G1024A rVP1, and F1022A/K1023A/G1024A rVP1 were electrophoresed in a 4-15% SDS-polyacrylamide gel and visualized by Coomassie Blue stain. A molecular mass (in kilodaltons) marker is shown to the left. (C) Radiolabeled dsRNA synthesis products synthesized by 2 pmol of each rVP1 in the presence of 8 pmol of rVP2 from strain SA11 rVP2. All reactions contained 8 pmol of an RVA +RNA template and were incubated at 37°C for 180 minutes. Radiolabeled dsRNA products were resolved in 4-15% SDS-polyacrylamide gels and visualized using a phosphorimager. (D) Radiolabeled dsRNA from 6 independent experiments using at least 3 protein batches was quantified and expressed as relative units (RUs). Averages are shown as bar graphs below each gel, and error bars represent standard deviations from the mean.

by VP2 (55). In this study, we employed an *in vitro* dsRNA synthesis assay and a loss-of-function mutagenesis approach to investigate whether these VP2 contact sites are important for polymerase activation. First, we used the available structural information for VP1 in the DLP to visualize five regions of the polymerase (i.e., EC1, PDC1, PDC2, EC2, and EC3) that were both surface-exposed and located on the side of VP1 that borders the VP2 core shell. Then, we used sequence analyses to assess the conservation of amino acids within these five contact sites to guide our mutagenesis studies. We posited that VP1 amino acid residues that (i) maintained a high degree of amino acid conservation and (ii) contained surface-exposed, electrically-charged side chains would be essential for VP2 interactions.

Interesting, rVP1 mutants with lesions in two of the five regions tested in these studies, specifically PDC1 and EC3, were fully functional and showed dsRNA synthetic activities that were comparable to WT rVP1. Residues within PDC1 were generally highly conserved across RV strains and contained mostly polar amino acids with one hydrophobic residue in the middle (**Fig 4-4A**). Although some amino acids within PDC1 have outward facing side chains that neighbor the VP2 core shell, they do not directly clash with any of the VP2 monomers in the DLP structure (**Fig. 4-4A**). Alanine point mutations at each individual amino acid residue yielded four single-point rVP1 mutants that maintained the capacity to synthesize dsRNA at levels that were indistinguishable from WT rVP1 (**Fig. 4-4C-D**). Given that PDC1 lacks electrically-charged residues and that mutating any one PDC1 residue is not sufficient to effect rVP1 functionality, it is unlikely that PDC1 is involved in mediating VP1-VP2 interactions that are critical for polymerase activation. Likewise, EC3, which contains four completely conserved amino acids including two positively charged lysine residues, is also not important for activation (**Fig. 4-7**). Within EC3 we made rVP1 mutants with two or three consecutive alanine residues,

including an obtrusive lysine to alanine substitution (**Fig. 4-7B**). Since several EC3 residues contained long amino acid side chains that extended toward the VP2 core shell and directly clash with VP2 in the DLP, we were surprised to observe that mutations within EC3 were not sufficient to effect rVP1 functionality in dsRNA synthesis assays (**Fig. 4-7C-D**). Thus, it is likely that EC3-VP2 contacts are not important for VP1 activation. However, our current results do not rule out the possibility that EC3-VP2 contacts could be important for mediating transcriptase VP1 functions, but not replicase functionality.

Our data revealed three VP1 regions (i.e., EC1, PDC2, and EC2) with surface-exposed residues important for VP2-dependent dsRNA synthesis. EC1 includes two consecutive-negatively charged glutamic acid residues within the region that directly overlaps with VP2 in the DLP (**Fig. 4-3A**). Interestingly, a drastic glutamic acid to alanine substitution within EC1 was only sufficient to reduce rVP1 functionality in the presence of an additional, albeit more modest, leucine-to-alanine mutation (i.e., E265A/L267A rVP1) (**Fig. 4-3C-D**). These results suggest that combinatorial effects from residues surrounding a prominent electrically-charged, outward facing amino acid are necessary to support interactions with VP2. In contrast, an R614A rVP1 single-point mutant within PDC2 showed significantly reduced dsRNA synthesis levels in comparison to WT, indicating that a similarly dramatic arginine to alanine substitution within PDC2 is sufficient to effect rVP1 functionality (**Fig. 4-5C-D**). Although PDC2 does not directly intersect VP2 in the context of the DLP, this region is proximally located to a loop-helix VP2 structure (aa 337-373) that protrudes downwards towards PDC2 (**Fig. 4-5A**). Based on these data, it is likely that diminished R614A rVP1 function can be attributed to inhibited VP2 contacts at this site. EC2, on the other hand, directly clashes with VP2 in the DLP (55). Though there are several electrically-charged amino acids within EC2, D971 was particularly interesting

because the negatively-charged aspartic acid is highly conserved in group A RV strains, and the amino acid side chain extends outward toward the VP2 core shell (**Fig. 4-6A**). We engineered three rVP1 mutants including an D971A mutation and 1-3 additional alanine substitutions for surrounding amino acids within EC2 (**Fig. 4-6B**). Expectedly, rVP1 proteins containing the D971A mutation synthesized significantly less dsRNA product (**Fig 4-6C**). We attempted to express and purify an rVP1 protein containing a single D971A mutation, but unfortunately this protein was not expressed at high enough levels to assay for replicase activity (data not shown). Nevertheless, these results support our original findings for EC1 and PDC2 that confirm the importance of highly conserved, electrically-charged and surface-exposed amino acid residues within EC2.

Structural analyses suggest that EC1, PDC2, and EC2 each interact with a distinct VP2 monomer to mediate dsRNA synthesis (**Fig. 4-8**). More specifically, EC1 is presumed to interact with VP2-A2 (UK VP2 aa 337-339: KEL), PDC2 is presumed to interact with VP2-B1 (UK VP2 aa I370), and EC2 is presumed to interact with VP2-A1 (UK VP2 aa L353, Q365, F366 and I370) (**Fig. 4-8**). Our results suggest that these interaction interfaces are critical for maintaining VP1 functionality, though it is possible that loss-of-function mutants described herein mediate dsRNA synthesis in a VP2-independent manner. Specifically, the *in vitro* RNA synthesis assays employed in this study do not explicitly measure VP2 binding, but rather detect an outcome of supposed interactions. Since rVP1 function is also template-dependent, it is possible that RNA-protein interactions mediate rVP1 function rather than protein-protein interactions. Future experiments in our lab will employ additional protein-RNA and protein-protein interaction assays to strengthen our conclusions that direct VP1-VP2 interactions are necessary for polymerase activation and that the engineered alanine mutations abrogated VP2 binding. Also,

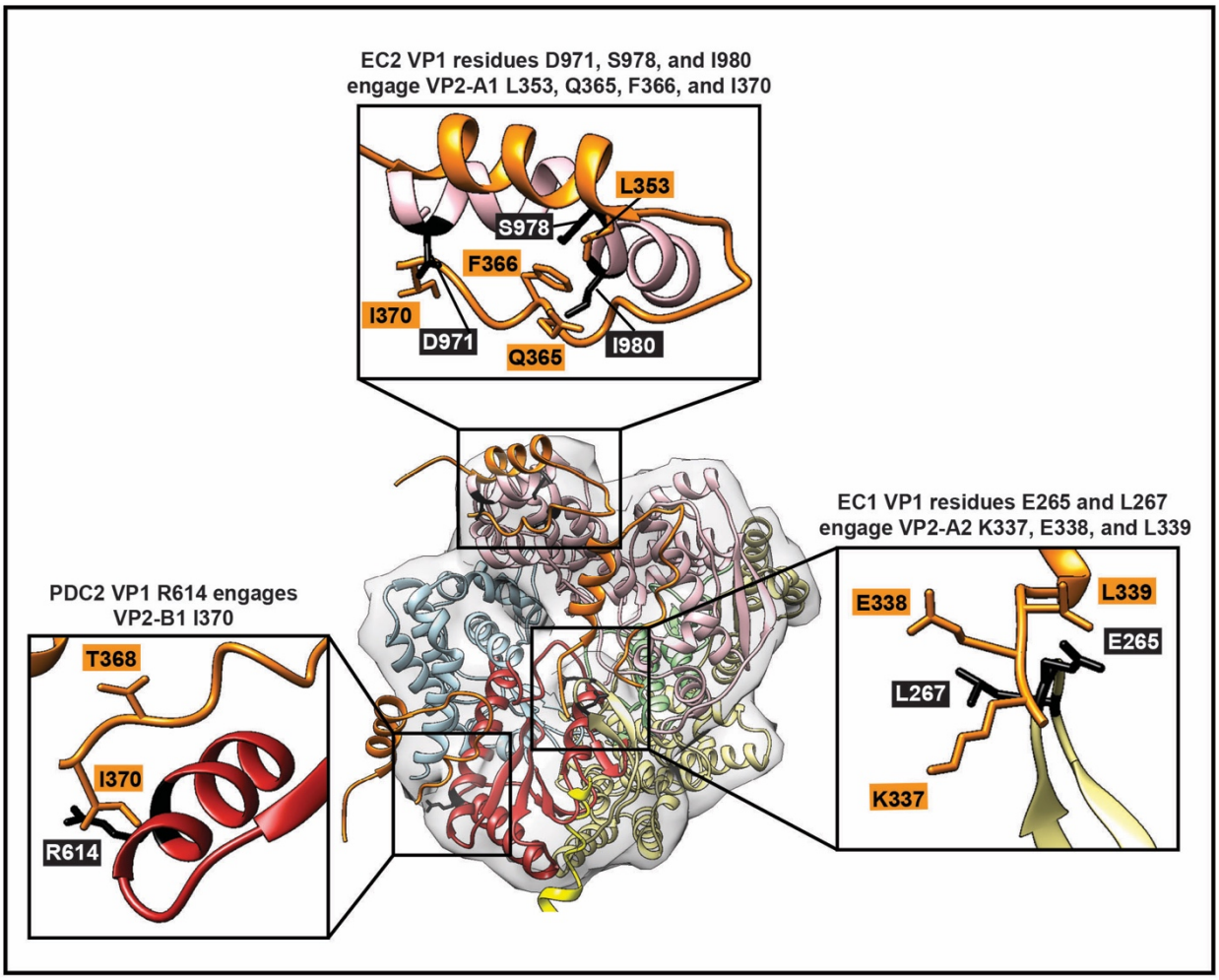


Fig. 4-8. Summary of VP1-VP2 contacts important for dsRNA synthesis. (A) A ribbon drawing of VP1 (PDB accession no. 4AU6 and 4F5X) is shown centrally in the same orientation and coloration as in Fig. 4-2C. Ribbon drawings of VP2 contact sites (UK VP2 aa 73-100 and 337-370) are shown in the in the same orientation and coloration as in Fig. 4-2E. Black boxes outline areas that are magnified in the central image. For clarity, some regions of VP1 and VP2 have been removed in the magnified images. Side chains are shown in black for regions EC1, PDC2, and EC2 and residues that showed loss of function when mutated to alanine are labeled. Side chains of corresponding regions of VP2 that may be interacting with regions EC1, PDC2, and EC2 are labeled and shown in orange. EC1 is presumed to interact with VP2-A2 (UK VP2 aa 337-339: KEL), PDC2 is presumed to interact with VP2-B1 (UK VP2 aa I370), and EC2 is presumed to interact with VP2-A1 (UK VP2 aa L353, Q365, F366 and I370).

at this time, we cannot exclude the possibility that a single point mutation could cause local conformational changes or intramolecular interactions within VP1 that inhibit polymerase activation independent of VP2 binding. Additional studies will benefit from more extensive structural analyses including molecular dynamics simulation of loss-of-function homology

models in order to compare predicted enzyme conformations and putative flexibility of mutated regions.

Future studies in our laboratory will also continue to assess these specific interaction interfaces using complementary mutational analyses of putative VP2 contact sites. Moreover, Estrozi and colleagues also identified two additional contact sites involving the amino-terminal domain of VP2 and regions of VP1 distal to the priming loop, which stabilizes the initiating nucleotide during RNA synthesis, and the catalytic center (55). The VP2 amino terminus domain (amino acids ~1-100) is predicted to protrude towards VP1 in the core, but remains incompletely resolved, with only a portion (amino acids ~80-100) resolved for VP2-B in the DLP structure (49, 55). Nonetheless, our results support a model in which VP2 engages VP1 at surface-exposed contact sites surrounding the +RNA exit tunnel, which triggers VP1 activation via an unbeknownst mechanism. Interestingly, we also uncovered amino acid signatures unique to avian RV strains, particularly within EC2, which may explicate previously reported strain-specific functionalities of rVP2-dependent rVP1 activity (119, 120). More specifically, previous data showed that rVP2-dependent rVP1 activation was inhibited by functionally incompatible avian and mammalian RV VP1 and VP2 proteins, but not within various mammalian RV strains (120). Thus, amino acid differences between avian and mammalian RV strains within EC2 could mediate interactions with VP2 that are critical for VP1 function, though additional experiments are required to validate these predictions.

Altogether, these data suggest that VP2-dependent VP1 function is mediated by some amino acid residues that overlap with VP2 contacts revealed from the DLP structure. More broadly, our results indicate that the polymerase may maintain its overall position and orientation against the VP2 core shell for both transcription and genome replication. Given that VP2-

dependent VP1 activation is a regulatory mechanism characteristic of both transcription and genome replication, it is possible that there are a single set of interacting VP1-VP2 residues that mediate VP1 activation during both transcription and genome replication. Thus, these challenges can be circumvented by using known structural data for the VP1-VP2 transcriptase complex to inform *in vitro* dsRNA synthesis assays.

Materials and Methods

Generation of recombinant VP1- and VP2-expressing baculoviruses. To create entry vectors for the generation of a baculovirus expressing single and multi-point rVP1 mutants, outward PCR and site directed mutagenesis was performed to modify the His-tagged pENTR-1A-SA11 VP1 clone generated by McDonald et al (183). In all PCR reactions Accuprime *Pfx* Supermix (Invitrogen) was used as the enzyme and the sequence was amplified using 5'-phosphorylated primers and pENTR-1A-SA11 VP1 as the template. Mutagenic primer sequences are available upon request. PCR products were treated with DpnI (New England BioLabs), and cDNAs were gel purified prior to ligation using T4 DNA ligase (New England BioLabs). To create an entry vector for the expression of SA11 rVP2, the cDNA was codon-optimized for expression in insect cells, synthesized *de novo* by GeneArt (Regensburg, Germany), and subcloned into pENTR-1A. The amino acid sequence of SA11 rVP2 protein directly matches the sequence in the GenBank database (accession no. DQ838631). All final pENTR-1A entry vector clones were all sequenced across the VP1/VP2 ORF prior to the generation of baculoviruses.

The BaculoDirect Expression System (Life Technologies) was used in accordance with the manufacturer's protocol to create recombinant baculoviruses expressing rVP1 or rVP2.

Briefly, the VP1- and VP2-encoding genes in the pENTR-1A entry vectors were individually inserted into BaculoDirect C-Term linear DNA by recombination with LR Clonase II. The baculovirus DNA was then transfected into *Spodoptera frugiperda* (Sf9) cells using Cellfectin reagent (Life Technologies) and recombinant baculovirus was harvested from selective medium containing 100 μM ganciclovir. Sf9 cells were maintained at 28°C in complete Grace's medium (Life Technologies) supplemented to contain 10% fetal bovine serum, 100 U/mL penicillin, 100 μg/mL streptomycin, 0.5 μg/mL amphotericin B, and 1% Pluronic F-68 (Life Technologies). Baculoviruses expressing wildtype SA11 rVP1 was generated previously (183).

Purification of rVP1 and rVP2 and *in vitro* dsRNA synthesis assay. His-tagged rVP1 and untagged rVP2 proteins were expressed and purified as described previously (119, 186). Purified rVP1 and rVP2 proteins were assessed for quality and relative quantity versus Precision Plus Protein Kaleidoscope Prestained Protein Standards (Bio-Rad) in 4-15% SDS-PAGE gels (Protean) and Coomassie Blue staining (Thermo Scientific). *In vitro* dsRNA synthesis assays were performed as described previously (118, 120). Briefly, each 20-μL reaction mixture contained 2 pmol rVP1; 8 pmol rVP2; 8 pmol SA11 gene 8 +RNA; 50 mM Tris-HCl (pH 7.5); 1 μL 30% polyethyleneglycol 8000; 20 mM magnesium acetate; 1.6 mM manganese acetate; 2.5 mM dithiothreitol; 1.25 mM (each) ATP, CTP, and UTP; 5 mM GTP; 1 μL RNasin (Promega); and 1 μCi [α-^{32}P]UTP (3,000 Ci/mmol; PerkinElmer). Reactions proceeded at 37°C for 3 hours. The [^{32}P]-labeled dsRNA products of the reaction were electrophoresed in 4-15% SDS-PAGE gels (Protean) and visualized using a GE Healthcare Storm 860 phosphorimager. A minimum of 3 replicate experiments were performed from 2-4 batches of independently purified recombinant proteins (n=6-12 replicates total). Quantification of images obtained using the phosphorimager was completed using ImageQuant 5.2 software or ImageJ 2.0v. One-sample *t* tests of the mean

were performed for each data set using Smith's Statistical Package 2.80v. *P* values of <0.005, in comparison to the set value of wildtype SA11 rVP1, were considered statistically significant.

Visual analyses of VP1 sequence variation and structure. Several representative full-length amino acid sequences were obtained from NCBI (www.ncbi.nlm.nih.gov/nuccore) including simian (si) strains SA11 (accession no. AFK09589) and RRV (accession no. ACC94312), human (hu) strains DS-1 (accession no. AEG25322), Wa (accession no. AFR77806), and Au-1 (accession no. ABF67540), bovine (bo) strains UK (accession no. P21615) and RF (P17468), murine (mu) strain ETD (accession no. AEA30040), and avian (av) strain PO-13 (accession no. BAA24146). Amino acid alignments were created using Geneious Pro v10.2.3 (BioMatters) and the BLOSUM matrix of the ClustalW algorithm. Structure images were generated using UCSF Chimera molecular modeling system v1.13 and Protein Data Bank files 4AU6 and 4F5X.

CHAPTER 5:

CONCLUSIONS AND DISCUSSION

Section 5.1 Summary of Results and Future Directions

RNA-dependent RNA polymerases (RdRps) are critical mediators of virus transcription and genome replication, yet they rarely act alone. The type, timing, and total amount of RNA synthesis is regulated by precise interactions between components of higher-order multisubunit polymerase complexes (182). Interactions with other cellular and/or viral proteins, nucleic acid, nucleotides, and ions, coordinate RNA synthesis with other events in the viral lifecycle and facilitate a productive viral infection. The rotavirus (RV) RdRp, VP1, is no exception. A myriad of co-factors is required for activation of VP1 and specific interactions among some of these co-factors may dictate whether VP1 functions as a transcriptase (i.e., mediates synthesis of +RNA) or a replicase (i.e., mediates synthesis of genomic dsRNA). Moreover, the VP2 core shell protein is an activator that triggers initiation of RNA synthesis during transcription and genome replication. High-resolution structures of the transcriptase complex have revealed a binding interface that overlaps regions surrounding the VP1 +RNA exit tunnel with the apical subdomain and N-terminal domain projections of VP2 (55). *In vitro* analyses confirmed that VP2 regions included in the binding footprint of the transcriptase complex were also involved in VP1 enzymatic activity during *in vitro* genome replication, indicating that these regions may be important for mediating RNA synthesis (119). In our studies, we sought to investigate the determinants of VP2-dependent VP1 activation using complementary bioinformatics and biochemical approaches. The results of these studies provide insight into how VP1 is functionally regulated by interactions with VP2 during dsRNA synthesis.

First, we sought to assess group A RV (RVA) strain diversity for VP1 and VP2 proteins, and consequently how the observed diversity affects VP2-dependent VP1 enzymatic function. Specifically, we used phylogenetic analyses and multi-dimensional scaling plots to compare VP1 and VP2 proteins encoded by genetically diverse RVA strains. We observed strong delineation of a subset of RVA strains that correlated with host species, such that VP1 and VP2 strains derived from avian hosts clustered distinctly from mammalian counterparts (120). To examine the extent to which VP1 and VP2 diversity is correlated with function, we assayed the functional compatibility of select recombinant (r) VP1 and VP2 proteins using an established *in vitro* dsRNA synthesis assay (118). We hypothesized that rVP2-dependent rVP1 activation would be maintained in reactions containing RVA strains from the same host type, but that host type-specific sequence factors would preclude rVP1 activity in dsRNA synthesis assays containing proteins derived from non-cognate host types (i.e., avian versus mammalian). Our results suggest that differences in rVP2-dependent rVP1 activation are not necessarily constrained by strain-specific sequence factors (120). However, additional studies are needed to completely characterize the capacity of divergent rVP1 and rVP2 proteins to functionally substitute. More comprehensive future studies should test the functional compatibility of additional representative RVA strains to expatiate on the functional differences between avian- versus mammalian-derived rVP1 and rVP2 proteins. However, such studies would require the development of a high-throughput functional assay, as the *in vitro* dsRNA synthesis assay described herein is time-intensive and laborious. Additionally, conclusions from the dsRNA synthesis assays are limited in scope, in that rVP2-dependent rVP1 function is assayed using an artificial and minimalistic *in vitro* replication system (118). Since the recent development of a robust fully-plasmid reverse genetics system (209), a more biologically significant experimental system would involve

synthesis of recombinant viruses containing VP1 and VP2 genes from varying RVA strains. However, recombinant viruses with incompatible VP1-VP2 proteins will most likely debilitate virus production; thus, it will be difficult to deduce whether absence of viral infection is due to inhibited RNA synthesis or a multitude of other factors, including segment mismatch, packaging inefficiency, or immune response. Therefore, an ideal research approach would involve the following pipeline: (i) comprehensive bioinformatic analyses, (ii) high-throughput assays to determine enzyme functionality, and (iii) analysis of recombinant virus phenotype using reverse genetics.

We next sought to determine which specific residues of VP1 are critical of VP2-dependent VP1 activation using site-directed mutagenesis and an *in vitro* dsRNA synthesis assay. First, we utilized these approaches in an attempt to engineer chimeric rVP1 mutants with a gain-of-function phenotype by exploiting functional incompatibilities between a mammalian-derived RVA strain rVP1 and an avian-derived RVA strain rVP2. Though we were not successful in synthesizing gain-of-function chimeric VP1 proteins, our results suggest that there are intricate intramolecular signaling pathways within rVP1 that are sensitive to amino acid changes, even at the protein surface, and that rVP2-dependent rVP1 activation may occur via an allosteric mechanism (120). Next, we made a series of single- and multi-point rVP1 mutants specifically targeting the VP1-VP2 binding interface identified in the structure of the transcriptase complex (55). Interestingly, the regions of VP1 and VP2 that comprise the binding footprint are highly conserved across divergent RVA strains, with the exception of one contact region, EC2 (SA11 VP1 aa 968-980), within the VP1 CTD (120). Thus, if VP1 function is mediated by interactions at the EC2 interface, then these observations indicate that (i) all VP2 proteins should functionally substitute for one another, and (ii) strain-specific functionality is

mediated by the CTD region of VP1. Our results do not show that all rVP2 proteins are functionally compatible (120). However, one single point mutant (R614A rVP1) and two multi point mutants (E265A/L267A rVP1 and D971A/S978A/I989A rVP1) showed loss of enzymatic function *in vitro*, which suggests that some amino acid residues within the transcriptase complex binding interface are important for RNA synthesis. These preliminary data that suggest that the interaction interface between VP1-VP2 during transcription may be similar during genome replication, though the absence of an *in vitro* transcription assays limits the strength of these conclusions. Given that these conclusions are derived from an *in vitro* functional assay, rather than a direct binding assay, future studies should aim to directly measure protein binding affinity and kinetics in order to elucidate this interaction(s) more completely. Studies investigating this interaction to date have been limited by sub-optimal purification of recombinant VP2, resulting in a population of heterogeneous multimers contaminated with proteolysis products, thereby precluding standard protein-protein interaction assays (46, 198, 200). Future studies would benefit from improved VP2 purification protocols, which could be achieved by optimizing buffer compositions or using an alternative expression system. Indeed, the multimerization status of the VP2 molecule that supports VP1 activation is not yet know and moreover, the molecular formation of VP2 may differ *in vitro* versus *in vivo.* As an alternative to direct protein binding assays, future investigators could employ a virus-like particle (VLP) approach whereby encapsidation of rVP1 into double-shelled particles comprised of rVP2 and rVP6 is indicative of rVP1-rVP2 binding interactions (46, 198, 201).

Section 5.2 Comparison of RdRp Structure and Mechanism

Even though VP1 contains all necessary catalytic motifs and has the capacity to bind template RNA, VP1 lacks polymerase activity in the absence of its activator protein, VP2. Given that diverse polymerases share common mechanisms of catalysis, conserved structural features, and are ubiquitously modulated by interactions with various co-factors, it is possible that the mechanisms by which VP2 interactions modulate VP1 function may be shared with other RNA viruses. For example, RNA and DNA polymerases alike catalyze phosphodiester bond reactions using a two-metal-ion mechanism of catalysis. In fact, two metal ions have consistently been found in the active sites of every DNA and RNA polymerase complexed with template and dNTP/NTPs (210-214). In this approach, two consecutive, highly conserved aspartate residues coordinate two metal ions (usually divalent magnesium cations) in the catalytic cleft. Both divalent cations are necessary for phosphodiester bond formation: one metal ion is important for formation of the nucleophile, whereas the other metal ion is involved in transition state stabilization (125, 127). VP1 maintains invariant Glu-Asp-Asp residues within its active site and these residues are directly implicated in catalysis, presumably via interactions with divalent cations in the catalytic pocket (207, 208). Additionally, polymerase structures remain strikingly conserved, especially across disparate groups of RNA viruses (182, 215). For example, RdRps of RV, *Pseudomonas* phage φ6, vesicular stomatitis virus, influenza virus, norovirus, and Zika virus all adopt a closed structure comprised of a canonical right-handed conformation with fingers, palm, and thumb subdomains (43, 216-220) (**Fig 5-1**). Though some structural variations are accommodated, mostly within the thumb subdomain and depending on mode of initiation of replication (e.g., primer-dependent versus *de novo* initiation), the subdomains are largely involved in template recognition and binding, selection and stabilization of NTPs and

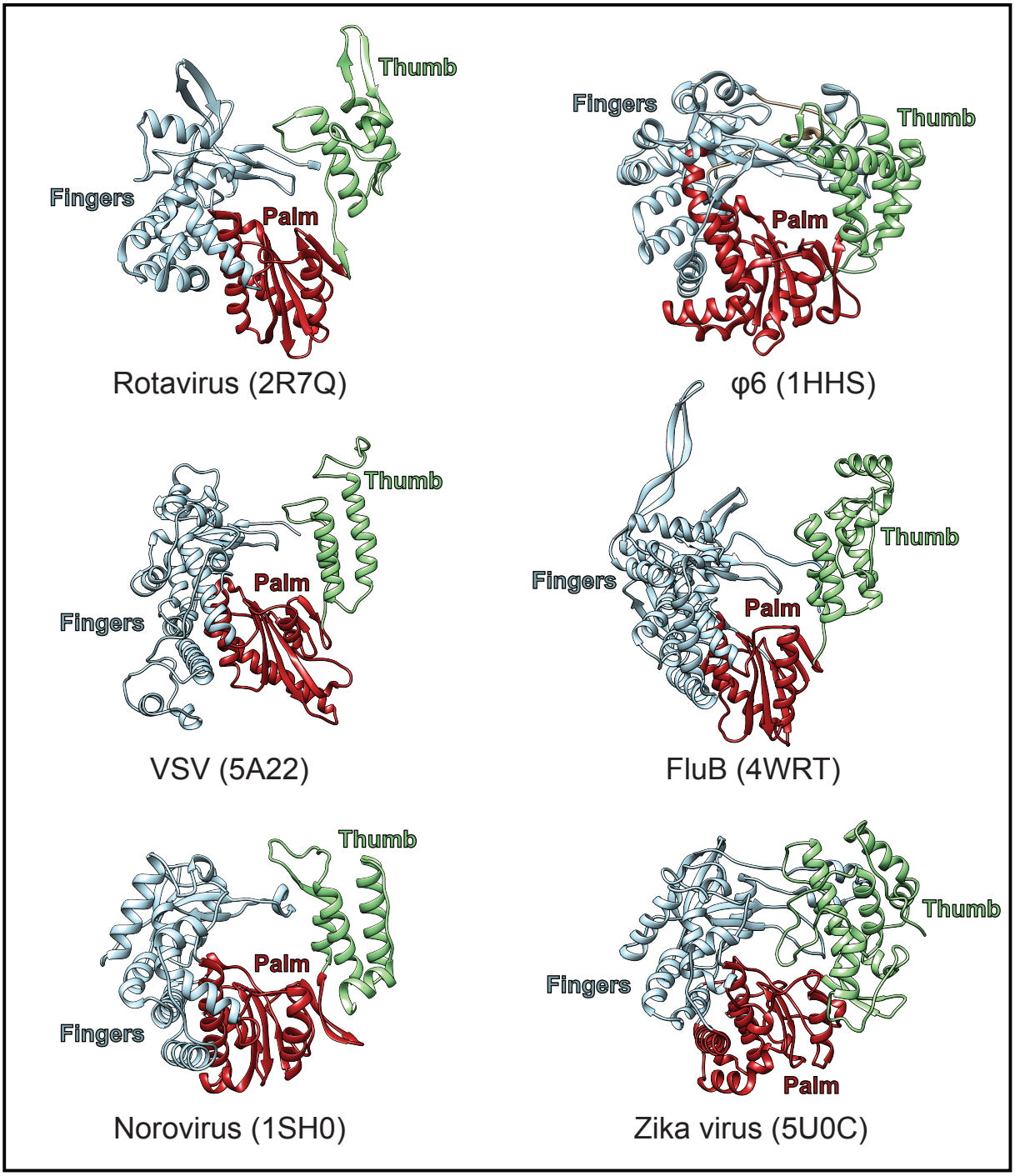


Fig. 5-1. Conserved three-dimensional architecture of RdRps from disparate RNA viruses. Ribbon representations of RdRp polymerase domains of: rotavirus (PDB accession no. 2R7Q, VP1 residues 333-778), *Pseudomonas* phage φ6 (PDB accession no. 1HHS, P2 residues 1-600), influenza B virus (PDB accession no. 4WRT, PB1 residues 36-616), vesicular stomatitis virus (PDB accession no. 5A22, L residues 360-865), norovirus RdRp (PDB accession no. 1SH0, residues 59-487), and zika virus RdRp (PDB accession no. 5U0C, NS5 residues 305-903). The fingers, palm, and thumb subdomains are shown in light blue, red, and light green, respectively.

metal ions, and facilitation of polymerization (221). Six conserved structural motifs (A-F) are also shared among RdRps and impart various functional roles during RNA synthesis, including Motif C which contains the Glu-Asp-Asp active site residues within the palm subdomain (222). Though VP1 contains two accessory domains (NTD and CTD) in addition to the central polymerase domain, many structural elements critical to catalysis are maintained, such as the formation of tunnels that extend into the catalytic core (43, 207). Moreover, high-resolution structures of particle-bound *Reoviridae* polymerase complexes have demonstrated analogous orientation and position of RdRps within a capsid (37, 38, 55).

Interestingly, compartmentalization of genome replication is not limited to *Reoviridae*, and, in fact, has emerged as a common theme for several disparate classes of RNA viruses and reverse transcribing viruses (223). For example, reverse-transcribing hepatitis B viruses also employ a capsid-dependent strategy of genome replication wherein pre-genomic RNA is selectively packaged into progeny capsids and polymerase activity of the co-packaged reverse transcriptase (P protein) is modulated by interactions with the capsid (224). Additionally, viruses have evolved other mechanisms for polymerase regulation. For instance, viruses that encode a +RNA genome typically employ a polyprotein expression strategy whereby activity of the RdRp is regulated by proteolysis. For instance, enzymatic activity of the poliovirus RdRp (3D) is dependent upon viral protein-induced cleavage from the 3C protein (225-227). Conversely, RNA synthesis by influenza A virus is dependent upon formation of a heterotrimeric polymerase complex including the RdRp, PB1, and two viral activator proteins, PB2 and PA, that are stably linked head-to-tail in the order PA-PB1-PB2 (228). Another commonly employed polymerase regulation strategy involves modulation of template binding. Members of the *Rhabdoviridae* family including rabies virus and vesicular stomatitis virus utilize this strategy

whereby the viral RdRp (L) does not directly engage the RNA template, but rather requires the viral phosphoprotein (P) as a co-factor 'bridge' (217, 229). Certainly, it is possible (and likely) that the RV has developed a polymerase regulatory mechanism that shares features with some of these strategies (**Fig. 5-2**). In combination with biochemical analysis, high-resolution structures and novel imaging technologies are providing unprecedented insight into complex structure-function relationships and will be instrumental in expatiating these interactions.

Though the mechanism by which VP1 and VP2 interact to coordinate RNA synthesis are not yet fully understood, VP2-dependent regulation of VP1 activation is advantageous in that the dsRNA RV genome will only be synthesized within a nascent particle that is a precursor to a mature virion. Thus, the dsRNA genome, which is a potent trigger of the host antiviral response, is concealed within a particle and protected from host cell degradation processes. All available data suggest that VP2 binds VP1 and triggers initiation of RNA synthesis (42, 118, 119), though direct VP1-VP2 binding has not been experimentally validated and it is not yet known whether VP1-VP2 interactions are transient (i.e., smack-release) or stable (e.g., reminiscent of the influenza A polymerase complex). The leading hypothesis in the field posits that VP2 engages a surface-exposed binding site on VP1, which prompts an intramolecular signaling cascade that results in conformational changes that facilitate RNA synthesis (**Fig 5-2A**) (58, 114, 182). This hypothesis is supported by analyses of the catalytically inactive VP1 apoenzyme that possesses several misaligned elements characteristic of a pre-initiation state (43). For example, analysis of complexes formed by VP1 and short oligoribonucleotide strands containing the conserved 3' terminus (i.e., 5'-UGUGACC-3') revealed that the +RNA strand binds one nucleotide out-of-register at the initiation site (43). Moreover, in these complexes the so-called VP1 priming loop exists in a retracted position, which inhibits binding of the first nucleotide (43). Given that VP1

becomes enzymatically active in the presence of VP2, it is proposed that interactions between VP1 and VP2 generate intramolecular structural rearrangements that cause 'in-register' positioning of the +RNA template in the template entry tunnel and extension of the priming loop, thereby supporting nucleotide binding in the priming site and allowing phosphodiester bonds to form. However, the dynamic capacity of VP1 is not well understood; VP2 binding to surface-

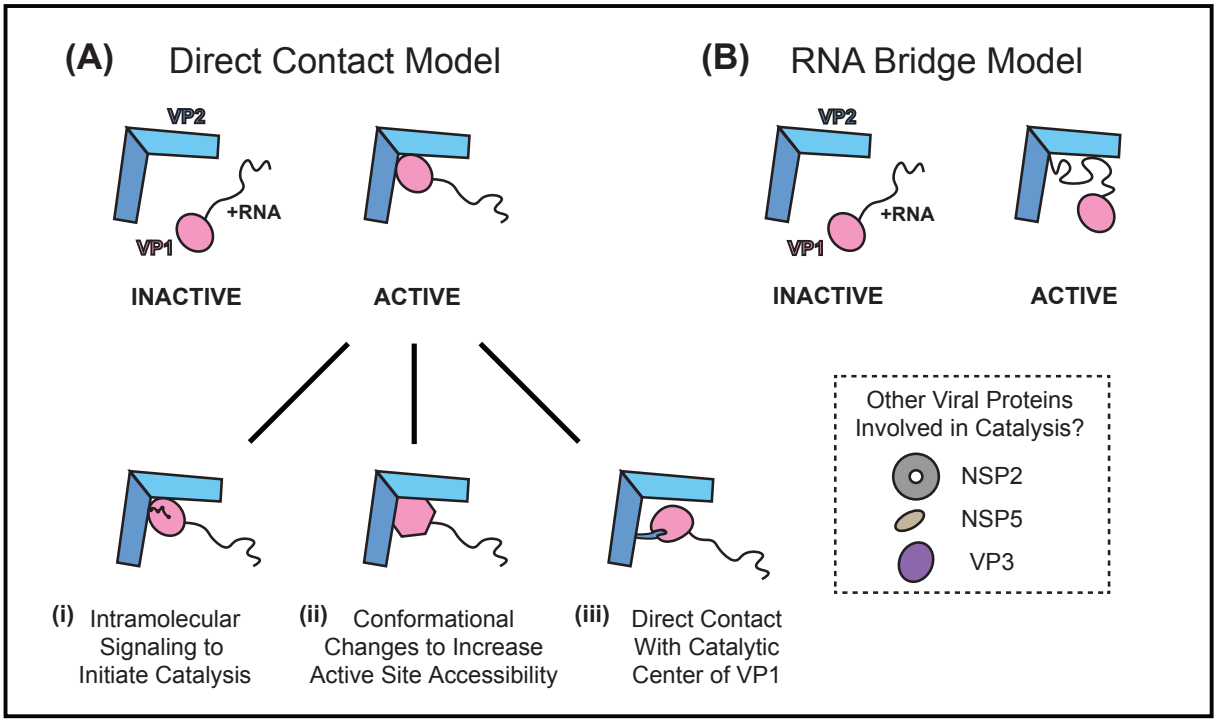


Fig. 5-2. Putative models of rotavirus polymerase activation. Several hypothetical models of VP1 activation are summarized in cartoon form. At minimum, enzymatic activity requires VP2 and/or +RNA to interact directly (A) or indirectly (B) with VP1. If VP1 and VP2 bind directly, several mechanisms could result trigger initiation of RNA synthesis. For example, VP2 interactions with surface-exposed VP1 residues may prompt an intramolecular signaling cascade that causes a series of conformational changes within the VP1 enzyme that culminate at the active site to initiate catalysis (i). Conversely, it is possible that VP2 binding at the VP1 surface causes conformational changes that increase accessibility of the active site, for example by widening the NTP entry tunnel to allow for initation of RNA synthesis (ii). In contrast, VP2 may interact directly with the catalytic center of VP1 via flexible N-terminal protrusions that may stimulate synthetic activities of VP1 (iii). It is also possible that VP1 and VP2 do not directly contact, but rather interact via an RNA "bridge" model wherein VP2 binds +RNA, which causes rearrangements to the secondary structure of +RNA and confers proper binding of +RNA to VP1 (B).

exposed residues of VP1 could cause internal signaling cascades that ultimately cumulate at the active site and trigger catalysis, or VP2 interactions with the VP1 surface may cause conformational shifts that increase accessibility of the active site, for example by expanding the diameter of the NTP entry tunnel (**Fig 5-2A**). Moreover, the structure, function, location, and orientation of the VP2-N terminal domain (NTD) is largely unidentified, but it is possible that the flexible N-terminal extensions of VP2 may protrude into the catalytic core and directly stimulate catalysis (**Fig 5-2A**). Additionally, current findings do not rule out the possibility that VP1 function is regulated by a mechanism independent of direct interaction with VP2, reminiscent of the RNA bridge model of VSV RdRp activation. For example, it is possible that VP2 binds the +RNA template, rather than VP1, and induces structural changes within the +RNA structure that facilitates binding to VP1 and induction of RNA synthesis (**Fig 5-2B**). Additional insights into VP1-VP2 interactions and mechanistic foundations of RNA initiation are dependent upon high-resolution structures for the VP1 polymerase complex in an enzymatically-active, post-initiation state (e.g., a co-crystal structure of VP1, VP2, and +RNA), though such studies have so far been hindered by challenges in obtaining highly pure, homogenous VP2 preparations.

Section 5.3 Significance of Results

Regardless of regulatory mechanism, nearly all RNA viruses are dependent upon functions of a virally-encoded RdRp to enable virus replication and establish a productive infection. Since RdRps are specialized enzymes that function to replicate viral RNA genomes, there are no cellular homologs present in an uninfected host. Since no functional equivalent to the RdRP exists in uninfected mammalian cells, RdRPs are a particularly attractive antiviral drug

target because it is likely that specific inhibitors can block viral replication with minimal associated toxicity and reduced side effects in comparison with corresponding orthosteric drugs (230). In fact, RdRps are the leading target for antiviral therapeutic drugs. For example, sofosbuvir (Sovaldi), a ribonucleoside analog, has been approved and has been successful in treating hepatitis C virus. Additionally, >13 drugs have targeted the HIV reverse transcriptase (231) and promising allosteric RdRP inhibitors have been identified for poliovirus (232), dengue virus (233), and norovirus (234). But, given that current RV vaccines are so effective, RV antiviral development is unlikely. It is more likely, however, that results from these studies may apply broadly to other RdRps, due to shared structural and functional features of RdRps encoded by divergent viruses.

While advances have been made in elucidating the roles of VP1 and VP2 in RV replication, the functional milieu of the active polymerase complex has yet to be fully revealed. Though the research described herein contributes important data to the RV field, many questions remain unanswered. What are the interaction interfaces required for catalysis? Is VP1 activated by VP2, or is VP1-VP2 the active polymerase complex? Is the mode of VP1-VP2 interaction the same during transcription and genome replication? If so, what triggers differentiation between single- versus double-stranded RNA synthesis? Additional studies investigating mechanisms of RV replication and RdRp biology are needed to describe these processes in detail. Indeed, such fundamental research will be essential in the quest to alleviate the effects of devastating viral diseases worldwide.

REFERENCES:

1. Liu L, Oza S, Hogan D, Perin J, Rudan I, Lawn JE, Cousens S, Mathers C, Black RE. 2015. Global, regional, and national causes of child mortality in 2000-13, with projections to inform post-2015 priorities: an updated systematic analysis. Lancet 385:430-40.
2. Lozano R, Naghavi M, Foreman K, Lim S, Shibuya K, Aboyans V, Abraham J, Adair T, Aggarwal R, Ahn SY, Alvarado M, Anderson HR, Anderson LM, Andrews KG, Atkinson C, Baddour LM, Barker-Collo S, Bartels DH, Bell ML, Benjamin EJ, Bennett D, Bhalla K, Bikbov B, Bin Abdulhak A, Birbeck G, Blyth F, Bolliger I, Boufous S, Bucello C, Burch M, Burney P, Carapetis J, Chen H, Chou D, Chugh SS, Coffeng LE, Colan SD, Colquhoun S, Colson KE, Condon J, Connor MD, Cooper LT, Corriere M, Cortinovis M, de Vaccaro KC, Couser W, Cowie BC, Criqui MH, Cross M, Dabhadkar KC, et al. 2012. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet 380:2095-128.
3. GBD Mortality Causes of Death, Collaborators. 2015. Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet 385:117-71.
4. World Health Organization. 2017. https://www.who.int/news-room/fact-sheets/detail/diarrhoeal-disease. Accessed 12 December 2018.
5. Bryce J, Boschi-Pinto C, Shibuya K, Black RE, Group WHOCHER. 2005. WHO estimates of the causes of death in children. Lancet 365:1147-52.

6. Tate JE, Burton AH, Boschi-Pinto C, Parashar UD, World Health Organization-Coordinated Global Rotavirus Surveillance N. 2016. Global, Regional, and National Estimates of Rotavirus Mortality in Children <5 Years of Age, 2000-2013. Clin Infect Dis 62 Suppl 2:S96-S105.
7. de Zoysa I, Feachem RG. 1985. Interventions for the control of diarrhoeal diseases among young children: rotavirus and cholera immunization. Bull World Health Organ 63:569-83.
8. Graham DY, Dufour GR, Estes MK. 1987. Minimal infective dose of rotavirus. Arch Virol 92:261-71.
9. Estes MK, Graham DY, Smith EM, Gerba CP. 1979. Rotavirus stability and inactivation. J Gen Virol 43:403-9.
10. Kapikian AZ, Wyatt RG, Levine MM, Yolken RH, VanKirk DH, Dolin R, Greenberg HB, Chanock RM. 1983. Oral administration of human rotavirus to volunteers: induction of illness and correlates of resistance. J Infect Dis 147:95-106.
11. Ward RL, Bernstein DI, Young EC, Sherwood JR, Knowlton DR, Schiff GM. 1986. Human rotavirus studies in volunteers: determination of infectious dose and serological response to infection. J Infect Dis 154:871-80.
12. Ramig RF. 2004. Pathogenesis of intestinal and systemic rotavirus infection. J Virol 78:10213-20.
13. Greenberg HB, Estes MK. 2009. Rotaviruses: from pathogenesis to vaccination. Gastroenterology 136:1939-51.
14. Ball JM, Tian P, Zeng CQ, Morris AP, Estes MK. 1996. Age-dependent diarrhea induced by a rotaviral nonstructural glycoprotein. Science 272:101-4.

15. Hagbom M, Istrate C, Engblom D, Karlsson T, Rodriguez-Diaz J, Buesa J, Taylor JA, Loitto VM, Magnusson KE, Ahlman H, Lundgren O, Svensson L. 2011. Rotavirus stimulates release of serotonin (5-HT) from human enterochromaffin cells and activates brain structures involved in nausea and vomiting. PLoS Pathog 7:e1002115.

16. Lundgren O, Peregrin AT, Persson K, Kordasti S, Uhnoo I, Svensson L. 2000. Role of the enteric nervous system in the fluid and electrolyte secretion of rotavirus diarrhea. Science 287:491-5.

17. Michelangeli F, Ruiz MC, del Castillo JR, Ludert JE, Liprandi F. 1991. Effect of rotavirus infection on intracellular calcium homeostasis in cultured cells. Virology 181:520-7.

18. Ruiz-Palacios GM, Perez-Schael I, Velazquez FR, Abate H, Breuer T, Clemens SC, Cheuvart B, Espinoza F, Gillard P, Innis BL, Cervantes Y, Linhares AC, Lopez P, Macias-Parra M, Ortega-Barria E, Richardson V, Rivera-Medina DM, Rivera L, Salinas B, Pavia-Ruz N, Salmeron J, Ruttimann R, Tinoco JC, Rubio P, Nunez E, Guerrero ML, Yarzabal JP, Damaso S, Tornieporth N, Saez-Llorens X, Vergara RF, Vesikari T, Bouckenooghe A, Clemens R, De Vos B, O'Ryan M, Human Rotavirus Vaccine Study G. 2006. Safety and efficacy of an attenuated vaccine against severe rotavirus gastroenteritis. N Engl J Med 354:11-22.

19. Vesikari T, Matson DO, Dennehy P, Van Damme P, Santosham M, Rodriguez Z, Dallas MJ, Heyse JF, Goveia MG, Black SB, Shinefield HR, Christie CD, Ylitalo S, Itzler RF, Coia ML, Onorato MT, Adeyi BA, Marshall GS, Gothefors L, Campens D, Karvonen A, Watt JP, O'Brien KL, DiNubile MJ, Clark HF, Boslego JW, Offit PA, Heaton PM,

Rotavirus E, Safety Trial Study T. 2006. Safety and efficacy of a pentavalent human-bovine (WC3) reassortant rotavirus vaccine. N Engl J Med 354:23-33.

20. Velasquez DE, Parashar UD, Jiang B. 2014. Strain diversity plays no major role in the varying efficacy of rotavirus vaccines: an overview. Infect Genet Evol 28:561-71.

21. Tate JE, Burton AH, Boschi-Pinto C, Steele AD, Duque J, Parashar UD, Network WH-cGRS. 2012. 2008 estimate of worldwide rotavirus-associated mortality in children younger than 5 years before the introduction of universal rotavirus vaccination programmes: a systematic review and meta-analysis. Lancet Infect Dis 12:136-41.

22. Arnold MM. 2018. Rotavirus vaccines: why continued investment in research is necessary. Curr Clin Microbiol Rep 5:73-81.

23. Desselberger U. 2017. Differences of Rotavirus Vaccine Effectiveness by Country: Likely Causes and Contributing Factors. Pathogens 6.

24. Flewett TH, Bryden AS, Davies H. 1973. Letter: Virus particles in gastroenteritis. Lancet 2:1497.

25. Kapikian AZ, Wyatt RG, Dolin R, Thornhill TS, Kalica AR, Chanock RM. 1972. Visualization by immune electron microscopy of a 27-nm particle associated with acute infectious nonbacterial gastroenteritis. J Virol 10:1075-81.

26. Adams WR, Kraft LM. 1963. Epizootic diarrhea of infant mice: indentification of the etiologic agent. Science 141:359-60.

27. Schnagl RD, Holmes IH. 1976. Characteristics of the genome of human infantile enteritis virus (Rotavirus). J Virol 19:267-70.

28. Davidson GP, Bishop RF, Townley RR, Holmes IH. 1975. Importance of a new virus in acute sporadic enteritis in children. Lancet 1:242-6.

29. Flewett TH, Bryden AS, Davies H, Woode GN, Bridger JC, Derrick JM. 1974. Relation between viruses from acute gastroenteritis of children and newborn calves. Lancet 2:61-3.

30. Matthijnssens J, Otto PH, Ciarlet M, Desselberger U, Van Ranst M, Johne R. 2012. VP6-sequence-based cutoff values as a criterion for rotavirus species demarcation. Arch Virol 157:1177-82.

31. Mihalov-Kovacs E, Gellert A, Marton S, Farkas SL, Feher E, Oldal M, Jakab F, Martella V, Banyai K. 2015. Candidate new rotavirus species in sheltered dogs, Hungary. Emerg Infect Dis 21:660-3.

32. Banyai K, Kemenesi G, Budinski I, Foldes F, Zana B, Marton S, Varga-Kugler R, Oldal M, Kurucz K, Jakab F. 2017. Candidate new rotavirus species in Schreiber's bats, Serbia. Infect Genet Evol 48:19-26.

33. Fischer TK, Gentsch JR. 2004. Rotavirus typing methods and algorithms. Rev Med Virol 14:71-82.

34. Matthijnssens J, Ciarlet M, Heiman E, Arijs I, Delbeke T, McDonald SM, Palombo EA, Iturriza-Gomara M, Maes P, Patton JT, Rahman M, Van Ranst M. 2008. Full genome-based classification of rotaviruses reveals a common origin between human Wa-Like and porcine rotavirus strains and human DS-1-like and bovine rotavirus strains. J Virol 82:3204-19.

35. Rotavirus Classification Working Group. 2018. https://rega.kuleuven.be/cev/viralmetagenomics/virus-classification/rcwg. Accessed 12 December 2018.

36. Prasad BV, Rothnagel R, Zeng CQ, Jakana J, Lawton JA, Chiu W, Estes MK. 1996. Visualization of ordered genomic RNA and localization of transcriptional complexes in rotavirus. Nature 382:471-3.

37. Ding K, Nguyen L, Zhou ZH. 2018. In Situ Structures of the Polymerase Complex and RNA Genome Show How Aquareovirus Transcription Machineries Respond to Uncoating. J Virol 92.

38. Zhang X, Ding K, Yu X, Chang W, Sun J, Zhou ZH. 2015. In situ structures of the segmented genome and RNA polymerase complex inside a dsRNA virus. Nature 527:531-4.

39. Gouet P, Diprose JM, Grimes JM, Malby R, Burroughs JN, Zientara S, Stuart DI, Mertens PP. 1999. The highly ordered double-stranded RNA genome of bluetongue virus revealed by crystallography. Cell 97:481-90.

40. Patton JT, Vasquez-Del Carpio R, Tortorici MA, Taraporewala ZF. 2007. Coupling of rotavirus genome replication and capsid assembly. Adv Virus Res 69:167-201.

41. Estes MK, Cohen J. 1989. Rotavirus gene structure and function. Microbiol Rev 53:410-49.

42. Tortorici MA, Broering TJ, Nibert ML, Patton JT. 2003. Template recognition and formation of initiation complexes by the replicase of a segmented double-stranded RNA virus. J Biol Chem 278:32673-82.

43. Lu X, McDonald SM, Tortorici MA, Tao YJ, Vasquez-Del Carpio R, Nibert ML, Patton JT, Harrison SC. 2008. Mechanism for coordinated RNA packaging and genome replication by rotavirus polymerase VP1. Structure 16:1678-88.

44. Ogden KM, Johne R, Patton JT. 2012. Rotavirus RNA polymerases resolve into two phylogenetically distinct classes that differ in their mechanism of template recognition. Virology 431:50-7.

45. Rainsford EW, McCrae MA. 2007. Characterization of the NSP6 protein product of rotavirus gene 11. Virus Res 130:193-201.

46. Lawton JA, Zeng CQ, Mukherjee SK, Cohen J, Estes MK, Prasad BV. 1997. Three-dimensional structural analysis of recombinant rotavirus-like particles with intact and amino-terminal-deleted VP2: implications for the architecture of the VP2 capsid layer. J Virol 71:7353-60.

47. Li Z, Baker ML, Jiang W, Estes MK, Prasad BV. 2009. Rotavirus architecture at subnanometer resolution. J Virol 83:1754-66.

48. Chen JZ, Settembre EC, Aoki ST, Zhang X, Bellamy AR, Dormitzer PR, Harrison SC, Grigorieff N. 2009. Molecular interactions in rotavirus assembly and uncoating seen by high-resolution cryo-EM. Proc Natl Acad Sci U S A 106:10644-8.

49. McClain B, Settembre E, Temple BR, Bellamy AR, Harrison SC. 2010. X-ray crystal structure of the rotavirus inner capsid particle at 3.8 A resolution. J Mol Biol 397:587-99.

50. Settembre EC, Chen JZ, Dormitzer PR, Grigorieff N, Harrison SC. 2011. Atomic model of an infectious rotavirus particle. EMBO J 30:408-16.

51. Zhang X, Settembre E, Xu C, Dormitzer PR, Bellamy R, Harrison SC, Grigorieff N. 2008. Near-atomic resolution using electron cryomicroscopy and single-particle reconstruction. Proc Natl Acad Sci U S A 105:1867-72.

52. Prasad BV, Wang GJ, Clerx JP, Chiu W. 1988. Three-dimensional structure of rotavirus. J Mol Biol 199:269-75.

53. Estes MK, Graham DY, Mason BB. 1981. Proteolytic enhancement of rotavirus infectivity: molecular mechanisms. J Virol 39:879-88.

54. Nair N, Feng N, Blum LK, Sanyal M, Ding S, Jiang B, Sen A, Morton JM, He XS, Robinson WH, Greenberg HB. 2017. VP4- and VP7-specific antibodies mediate heterotypic immunity to rotavirus in humans. Sci Transl Med 9.

55. Estrozi LF, Settembre EC, Goret G, McClain B, Zhang X, Chen JZ, Grigorieff N, Harrison SC. 2013. Location of the dsRNA-dependent polymerase, VP1, in rotavirus particles. J Mol Biol 425:124-32.

56. Brandmann T, Jinek M. 2015. Crystal structure of the C-terminal 2',5'-phosphodiesterase domain of group A rotavirus protein VP3. Proteins 83:997-1002.

57. Ogden KM, Hu L, Jha BK, Sankaran B, Weiss SR, Silverman RH, Patton JT, Prasad BV. 2015. Structural basis for 2'-5'-oligoadenylate binding and enzyme activity of a viral RNase L antagonist. J Virol 89:6633-45.

58. Gridley CL, Patton JT. 2014. Regulation of rotavirus polymerase activity by inner capsid proteins. Curr Opin Virol 9:31-8.

59. Ciarlet M, Crawford SE, Estes MK. 2001. Differential infection of polarized epithelial cell lines by sialic acid-dependent and sialic acid-independent rotavirus strains. J Virol 75:11834-50.

60. Dormitzer PR, Sun ZY, Wagner G, Harrison SC. 2002. The rhesus rotavirus VP4 sialic acid binding domain has a galectin fold with a novel carbohydrate binding site. EMBO J 21:885-97.

61. Spence L, Fauvel M, Petro R, Bloch S. 1976. Haemagglutinin from Rotavirus. Lancet 2:1023.

62. Bastardo JW, Holmes IH. 1980. Attachment of SA-11 rotavirus to erythrocyte receptors. Infect Immun 29:1134-40.

63. Fiore L, Greenberg HB, Mackow ER. 1991. The VP8 fragment of VP4 is the rhesus rotavirus hemagglutinin. Virology 181:553-63.

64. Ciarlet M, Estes MK. 1999. Human and most animal rotavirus strains do not require the presence of sialic acid on the cell surface for efficient infectivity. J Gen Virol 80 (Pt 4):943-8.

65. Isa P, Arias CF, Lopez S. 2006. Role of sialic acids in rotavirus infection. Glycoconj J 23:27-37.

66. Lopez S, Arias CF. 2006. Early steps in rotavirus cell entry. Curr Top Microbiol Immunol 309:39-66.

67. Hu L, Crawford SE, Czako R, Cortes-Penfield NW, Smith DF, Le Pendu J, Estes MK, Prasad BV. 2012. Cell attachment protein VP8* of a human rotavirus specifically interacts with A-type histo-blood group antigen. Nature 485:256-9.

68. Ramani S, Hu L, Venkataram Prasad BV, Estes MK. 2016. Diversity in Rotavirus-Host Glycan Interactions: A "Sweet" Spectrum. Cell Mol Gastroenterol Hepatol 2:263-273.

69. Isa P, Sanchez-Aleman MA, Lopez S, Arias CF. 2009. Dissecting the role of integrin subunits alpha 2 and beta 3 in rotavirus cell entry by RNA silencing. Virus Res 145:251-9.

70. Graham KL, Takada Y, Coulson BS. 2006. Rotavirus spike protein VP5* binds alpha2beta1 integrin on the cell surface and competes with virus for cell binding and infectivity. J Gen Virol 87:1275-83.

71. Zarate S, Cuadras MA, Espinosa R, Romero P, Juarez KO, Camacho-Nuez M, Arias CF, Lopez S. 2003. Interaction of rotaviruses with Hsc70 during cell entry is mediated by VP5. J Virol 77:7254-60.

72. Graham KL, Halasz P, Tan Y, Hewish MJ, Takada Y, Mackow ER, Robinson MK, Coulson BS. 2003. Integrin-using rotaviruses bind alpha2beta1 integrin alpha2 I domain via VP4 DGE sequence and recognize alphaXbeta2 and alphaVbeta3 by using VP7 during cell entry. J Virol 77:9969-78.

73. Guerrero CA, Bouyssounade D, Zarate S, Isa P, Lopez T, Espinosa R, Romero P, Mendez E, Lopez S, Arias CF. 2002. Heat shock cognate protein 70 is involved in rotavirus cell entry. J Virol 76:4096-102.

74. Torres-Flores JM, Silva-Ayala D, Espinoza MA, Lopez S, Arias CF. 2015. The tight junction protein JAM-A functions as coreceptor for rotavirus entry into MA104 cells. Virology 475:172-8.

75. Arias CF, Silva-Ayala D, Lopez S. 2015. Rotavirus entry: a deep journey into the cell with several exits. J Virol 89:890-3.

76. Zou WY, Blutt SE, Crawford SE, Ettayebi K, Zeng XL, Saxena K, Ramani S, Karandikar UC, Zachos NC, Estes MK. 2017. Human Intestinal Enteroids: New Models to Study Gastrointestinal Virus Infections. Methods Mol Biol doi:10.1007/7651_2017_1.

77. Ramani S, Crawford SE, Blutt SE, Estes MK. 2018. Human organoid cultures: transformative new tools for human virus studies. Curr Opin Virol 29:79-86.

78. Gutierrez M, Isa P, Sanchez-San Martin C, Perez-Vargas J, Espinosa R, Arias CF, Lopez S. 2010. Different rotavirus strains enter MA104 cells through different endocytic pathways: the role of clathrin-mediated endocytosis. J Virol 84:9161-9.

79. Kim IS, Trask SD, Babyonyshev M, Dormitzer PR, Harrison SC. 2010. Effect of mutations in VP5 hydrophobic loops on rotavirus cell entry. J Virol 84:6200-7.

80. Diaz-Salinas MA, Romero P, Espinosa R, Hoshino Y, Lopez S, Arias CF. 2013. The spike protein VP4 defines the endocytic pathway used by rotavirus to enter MA104 cells. J Virol 87:1658-63.

81. Ludert JE, Michelangeli F, Gil F, Liprandi F, Esparza J. 1987. Penetration and uncoating of rotaviruses in cultured cells. Intervirology 27:95-101.

82. Ruiz MC, Abad MJ, Charpilienne A, Cohen J, Michelangeli F. 1997. Cell lines susceptible to infection are permeabilized by cleaved and solubilized outer layer proteins of rotavirus. J Gen Virol 78 (Pt 11):2883-93.

83. Abdelhakim AH, Salgado EN, Fu X, Pasham M, Nicastro D, Kirchhausen T, Harrison SC. 2014. Structural correlates of rotavirus cell entry. PLoS Pathog 10:e1004355.

84. Libersou S, Siebert X, Ouldali M, Estrozi LF, Navaza J, Charpilienne A, Garnier P, Poncet D, Lepault J. 2008. Geometric mismatches within the concentric layers of rotavirus particles: a potential regulatory switch of viral particle transcription activity. J Virol 82:2844-52.

85. Stacy-Phipps S, Patton JT. 1987. Synthesis of plus- and minus-strand RNA in rotavirus-infected cells. J Virol 61:3479-84.

86. Ayala-Breton C, Arias M, Espinosa R, Romero P, Arias CF, Lopez S. 2009. Analysis of the kinetics of transcription and replication of the rotavirus genome by RNA interference. J Virol 83:8819-31.

87. Patton JT, Gallegos CO. 1990. Rotavirus RNA replication: single-stranded RNA extends from the replicase particle. J Gen Virol 71 (Pt 5):1087-94.

88. Tortorici MA, Shapiro BA, Patton JT. 2006. A base-specific recognition signal in the 5' consensus sequence of rotavirus plus-strand RNAs promotes replication of the double-stranded RNA genome segments. RNA 12:133-46.

89. Piron M, Vende P, Cohen J, Poncet D. 1998. Rotavirus RNA-binding protein NSP3 interacts with eIF4GI and evicts the poly(A) binding protein from eIF4F. EMBO J 17:5811-21.

90. Gratia M, Sarot E, Vende P, Charpilienne A, Baron CH, Duarte M, Pyronnet S, Poncet D. 2015. Rotavirus NSP3 Is a Translational Surrogate of the Poly(A) Binding Protein-Poly(A) Complex. J Virol 89:8773-82.

91. Gonzalez RA, Espinosa R, Romero P, Lopez S, Arias CF. 2000. Relative localization of viroplasmic and endoplasmic reticulum-resident rotavirus proteins in infected cells. Arch Virol 145:1963-73.

92. Silvestri LS, Tortorici MA, Vasquez-Del Carpio R, Patton JT. 2005. Rotavirus glycoprotein NSP4 is a modulator of viral transcription in the infected cell. J Virol 79:15165-74.

93. Carreno-Torres JJ, Gutierrez M, Arias CF, Lopez S, Isa P. 2010. Characterization of viroplasm formation during the early stages of rotavirus infection. Virol J 7:350.

94. Cabral-Romero C, Padilla-Noriega L. 2006. Association of rotavirus viroplasms with microtubules through NSP2 and NSP5. Mem Inst Oswaldo Cruz 101:603-11.

95. Cheung W, Gill M, Esposito A, Kaminski CF, Courousse N, Chwetzoff S, Trugnan G, Keshavan N, Lever A, Desselberger U. 2010. Rotaviruses associate with cellular lipid droplet components to replicate in viroplasms, and compounds disrupting or blocking lipid droplets inhibit viroplasm formation and viral replication. J Virol 84:6782-98.

96. Contin R, Arnoldi F, Mano M, Burrone OR. 2011. Rotavirus replication requires a functional proteasome for effective assembly of viroplasms. J Virol 85:2781-92.

97. Silvestri LS, Taraporewala ZF, Patton JT. 2004. Rotavirus replication: plus-sense templates for double-stranded RNA synthesis are made in viroplasms. J Virol 78:7763-74.

98. Kumar M, Jayaram H, Vasquez-Del Carpio R, Jiang X, Taraporewala ZF, Jacobson RH, Patton JT, Prasad BV. 2007. Crystallographic and biochemical analysis of rotavirus NSP2 with nucleotides reveals a nucleoside diphosphate kinase-like activity. J Virol 81:12272-84.

99. Vasquez-Del Carpio R, Gonzalez-Nilo FD, Riadi G, Taraporewala ZF, Patton JT. 2006. Histidine triad-like motif of the rotavirus NSP2 octamer mediates both RTPase and NTPase activities. J Mol Biol 362:539-54.

100. Taraporewala Z, Chen D, Patton JT. 1999. Multimers formed by the rotavirus nonstructural protein NSP2 bind to RNA and have nucleoside triphosphatase activity. J Virol 73:9934-43.

101. Taraporewala ZF, Patton JT. 2001. Identification and characterization of the helix-destabilizing activity of rotavirus nonstructural protein NSP2. J Virol 75:4519-27.

102. Jiang X, Jayaram H, Kumar M, Ludtke SJ, Estes MK, Prasad BV. 2006. Cryoelectron microscopy structures of rotavirus NSP2-NSP5 and NSP2-RNA complexes: implications for genome replication. J Virol 80:10829-35.

103. Martin D, Ouldali M, Menetrey J, Poncet D. 2011. Structural organisation of the rotavirus nonstructural protein NSP5. J Mol Biol 413:209-21.

104. Viskovska M, Anish R, Hu L, Chow DC, Hurwitz AM, Brown NG, Palzkill T, Estes MK, Prasad BV. 2014. Probing the sites of interactions of rotaviral proteins involved in replication. J Virol 88:12866-81.

105. Lawton JA, Estes MK, Prasad BV. 2000. Mechanism of genome transcription in segmented dsRNA viruses. Adv Virus Res 55:185-229.

106. Gallegos CO, Patton JT. 1989. Characterization of rotavirus replication intermediates: a model for the assembly of single-shelled particles. Virology 172:616-27.

107. Borodavka A, Dykeman EC, Schrimpf W, Lamb DC. 2017. Protein-mediated RNA folding governs sequence-specific interactions between rotavirus genome segments. Elife 6.

108. Suzuki Y. 2015. A candidate packaging signal of human rotavirus differentiating Wa-like and DS-1-like genomic constellations. Microbiol Immunol 59:567-71.

109. McDonald SM, Nelson MI, Turner PE, Patton JT. 2016. Reassortment in segmented RNA viruses: mechanisms and outcomes. Nat Rev Microbiol 14:448-60.

110. Taniguchi K, Kojima K, Urasawa S. 1996. Nondefective rotavirus mutants with an NSP1 gene which has a deletion of 500 nucleotides, including a cysteine-rich zinc finger motif-encoding region (nucleotides 156 to 248), or which has a nonsense codon at nucleotides 153-155. J Virol 70:4125-30.

111. Patton JT, Taraporewala Z, Chen D, Chizhikov V, Jones M, Elhelu A, Collins M, Kearney K, Wagner M, Hoshino Y, Gouvea V. 2001. Effect of intragenic rearrangement and changes in the 3' consensus sequence on NSP1 expression and rotavirus replication. J Virol 75:2076-86.

112. Komoto S, Fukuda S, Ide T, Ito N, Sugiyama M, Yoshikawa T, Murata T, Taniguchi K. 2018. Generation of Recombinant Rotaviruses Expressing Fluorescent Proteins by Using an Optimized Reverse Genetics System. J Virol 92.

113. Fajardo T, Jr., Sung PY, Roy P. 2015. Disruption of Specific RNA-RNA Interactions in a Double-Stranded RNA Virus Inhibits Genome Packaging and Virus Infectivity. PLoS Pathog 11:e1005321.

114. Long CP, McDonald SM. 2017. Rotavirus genome replication: Some assembly required. PLoS Pathog 13:e1006242.

115. Patton JT, Gallegos CO. 1988. Structure and protein composition of the rotavirus replicase particle. Virology 166:358-65.

116. Boudreaux CE, Kelly DF, McDonald SM. 2015. Electron microscopic analysis of rotavirus assembly-replication intermediates. Virology 477:32-41.

117. Zhang X, Walker SB, Chipman PR, Nibert ML, Baker TS. 2003. Reovirus polymerase lambda 3 localized by cryo-electron microscopy of virions at a resolution of 7.6 A. Nat Struct Biol 10:1011-8.

118. Patton JT, Jones MT, Kalbach AN, He YW, Xiaobo J. 1997. Rotavirus RNA polymerase requires the core shell protein to synthesize the double-stranded RNA genome. J Virol 71:9618-26.

119. McDonald SM, Patton JT. 2011. Rotavirus VP2 core shell regions critical for viral polymerase activation. J Virol 85:3095-105.

120. Steger CL, Boudreaux CE, LaConte LE, Pease JB, McDonald SM. 2019. Group A Rotavirus VP1 Polymerase and VP2 Core Shell Proteins: Intergenotypic Sequence Variation and In Vitro Functional Compatibility. J Virol 93.

121. Chen D, Zeng CQ, Wentz MJ, Gorziglia M, Estes MK, Ramig RF. 1994. Template-dependent, in vitro replication of rotavirus RNA. J Virol 68:7030-9.

122. Chen D, Patton JT. 1998. Rotavirus RNA replication requires a single-stranded 3' end for efficient minus-strand synthesis. J Virol 72:7387-96.

123. Chen D, Patton JT. 2000. De novo synthesis of minus strand RNA by the rotavirus RNA polymerase in a cell-free system involves a novel mechanism of initiation. RNA 6:1455-67.

124. Chen D, Barros M, Spencer E, Patton JT. 2001. Features of the 3'-consensus sequence of rotavirus mRNAs critical to minus strand synthesis. Virology 282:221-9.

125. Steitz TA, Steitz JA. 1993. A general two-metal-ion mechanism for catalytic RNA. Proc Natl Acad Sci U S A 90:6498-502.

126. Steitz TA. 1999. DNA polymerases: structural diversity and common mechanisms. J Biol Chem 274:17395-8.

127. Yang W, Lee JY, Nowotny M. 2006. Making and breaking nucleic acids: two-Mg2+-ion catalysis and substrate specificity. Mol Cell 22:5-13.

128. Trask SD, Dormitzer PR. 2006. Assembly of highly infectious rotavirus particles recoated with recombinant outer capsid proteins. J Virol 80:11293-304.

129. O'Brien JA, Taylor JA, Bellamy AR. 2000. Probing the structure of rotavirus NSP4: a short sequence at the extreme C terminus mediates binding to the inner capsid particle. J Virol 74:5388-94.

130. Au KS, Mattion NM, Estes MK. 1993. A subviral particle binding domain on the rotavirus nonstructural glycoprotein NS28. Virology 194:665-73.

131. Taylor JA, O'Brien JA, Lord VJ, Meyer JC, Bellamy AR. 1993. The RER-localized rotavirus intracellular receptor: a truncated purified soluble form is multivalent and binds virus particles. Virology 194:807-14.

132. Tian P, Hu Y, Schilling WP, Lindsay DA, Eiden J, Estes MK. 1994. The nonstructural glycoprotein of rotavirus affects intracellular calcium levels. J Virol 68:251-7.

133. Hyser JM, Utama B, Crawford SE, Broughman JR, Estes MK. 2013. Activation of the endoplasmic reticulum calcium sensor STIM1 and store-operated calcium entry by rotavirus requires NSP4 viroporin activity. J Virol 87:13579-88.

134. Crawford SE, Hyser JM, Utama B, Estes MK. 2012. Autophagy hijacked through viroporin-activated calcium/calmodulin-dependent kinase kinase-beta signaling is required for rotavirus replication. Proc Natl Acad Sci U S A 109:E3405-13.

135. Crawford SE, Estes MK. 2013. Viroporin-mediated calcium-activated autophagy. Autophagy 9:797-8.

136. McNulty MS, Curran WL, McFerran JB. 1976. The morphogenesis of a cytopathic bovine rotavirus in Madin-Darby bovine kidney cells. J Gen Virol 33:503-8.

137. Musalem C, Espejo RT. 1985. Release of progeny virus from cells infected with simian rotavirus SA11. J Gen Virol 66 (Pt 12):2715-24.

138. Jourdan N, Maurice M, Delautier D, Quero AM, Servin AL, Trugnan G. 1997. Rotavirus is released from the apical surface of cultured human intestinal cells through nonconventional vesicular transport that bypasses the Golgi apparatus. J Virol 71:8268-78.

139. O'Neill LA, Bowie AG. 2010. Sensing and signaling in antiviral innate immunity. Curr Biol 20:R328-33.

140. Wilkins C, Gale M, Jr. 2010. Recognition of viruses by cytoplasmic sensors. Curr Opin Immunol 22:41-7.

141. Yoneyama M, Fujita T. 2010. Recognition of viral nucleic acids in innate immunity. Rev Med Virol 20:4-22.

142. Takeuchi O, Akira S. 2010. Pattern recognition receptors and inflammation. Cell 140:805-20.

143. Sen A, Pruijssers AJ, Dermody TS, Garcia-Sastre A, Greenberg HB. 2011. The early interferon response to rotavirus is regulated by PKR and depends on MAVS/IPS-1, RIG-I, MDA-5, and IRF3. J Virol 85:3717-32.

144. Broquet AH, Hirata Y, McAllister CS, Kagnoff MF. 2011. RIG-I/MDA5/MAVS are required to signal a protective IFN response in rotavirus-infected intestinal epithelium. J Immunol 186:1618-26.

145. Kato H, Takeuchi O, Sato S, Yoneyama M, Yamamoto M, Matsui K, Uematsu S, Jung A, Kawai T, Ishii KJ, Yamaguchi O, Otsu K, Tsujimura T, Koh CS, Reis e Sousa C, Matsuura Y, Fujita T, Akira S. 2006. Differential roles of MDA5 and RIG-I helicases in the recognition of RNA viruses. Nature 441:101-5.

146. Stetson DB, Medzhitov R. 2006. Type I interferons in host defense. Immunity 25:373-81.

147. Wietek C, O'Neill LA. 2007. Diversity and regulation in the NF-kappaB system. Trends Biochem Sci 32:311-9.

148. Honda K, Taniguchi T. 2006. IRFs: master regulators of signalling by Toll-like receptors and cytosolic pattern-recognition receptors. Nat Rev Immunol 6:644-58.

149. Silverman RH. 2007. Viral encounters with 2',5'-oligoadenylate synthetase and RNase L during the interferon antiviral response. J Virol 81:12720-9.

150. Barro M, Patton JT. 2005. Rotavirus nonstructural protein 1 subverts innate immune response by inducing degradation of IFN regulatory factor 3. Proc Natl Acad Sci U S A 102:4114-9.

151. Graff JW, Mitzel DN, Weisend CM, Flenniken ML, Hardy ME. 2002. Interferon regulatory factor 3 is a cellular partner of rotavirus NSP1. J Virol 76:9545-50.

152. Silverman RH, Weiss SR. 2014. Viral phosphodiesterases that antagonize double-stranded RNA signaling to RNase L by degrading 2-5A. J Interferon Cytokine Res 34:455-63.

153. Pott J, Stockinger S, Torow N, Smoczek A, Lindner C, McInerney G, Backhed F, Baumann U, Pabst O, Bleich A, Hornef MW. 2012. Age-dependent TLR3 expression of the intestinal epithelium contributes to rotavirus susceptibility. PLoS Pathog 8:e1002670.

154. Deal EM, Jaimes MC, Crawford SE, Estes MK, Greenberg HB. 2010. Rotavirus structural proteins and dsRNA are required for the human primary plasmacytoid dendritic cell IFNalpha response. PLoS Pathog 6:e1000931.

155. Desselberger U, Huppertz HI. 2011. Immune responses to rotavirus infection and vaccination and associated correlates of protection. J Infect Dis 203:188-95.

156. Offit PA, Blavat G. 1986. Identification of the two rotavirus genes determining neutralization specificities. J Virol 57:376-8.

157. Offit PA, Blavat G, Greenberg HB, Clark HF. 1986. Molecular basis of rotavirus virulence: role of gene segment 4. J Virol 57:46-9.

158. Aiyegbo MS, Sapparapu G, Spiller BW, Eli IM, Williams DR, Kim R, Lee DE, Liu T, Li S, Woods VL, Jr., Nannemann DP, Meiler J, Stewart PL, Crowe JE, Jr. 2013. Human

rotavirus VP6-specific antibodies mediate intracellular neutralization by binding to a quaternary structure in the transcriptional pore. PLoS One 8:e61101.

159. Aoki ST, Settembre EC, Trask SD, Greenberg HB, Harrison SC, Dormitzer PR. 2009. Structure of rotavirus outer-layer protein VP7 bound with a neutralizing Fab. Science 324:1444-7.

160. Taniguchi K, Urasawa T, Kobayashi N, Ahmed MU, Adachi N, Chiba S, Urasawa S. 1991. Antibody response to serotype-specific and cross-reactive neutralization epitopes on VP4 and VP7 after rotavirus infection or vaccination. J Clin Microbiol 29:483-7.

161. Padilla-Noriega L, Dunn SJ, Lopez S, Greenberg HB, Arias CF. 1995. Identification of two independent neutralization domains on the VP4 trypsin cleavage products VP5* and VP8* of human rotavirus ST3. Virology 206:148-54.

162. Mackow ER, Shaw RD, Matsui SM, Vo PT, Benfield DA, Greenberg HB. 1988. Characterization of homotypic and heterotypic VP7 neutralization sites of rhesus rotavirus. Virology 165:511-7.

163. Giammarioli AM, Mackow ER, Fiore L, Greenberg HB, Ruggeri FM. 1996. Production and characterization of murine IgA monoclonal antibodies to the surface antigens of rhesus rotavirus. Virology 225:97-110.

164. Jiang B, Gentsch JR, Glass RI. 2002. The role of serum antibodies in the protection against rotavirus disease: an overview. Clin Infect Dis 34:1351-61.

165. Coulson BS, Grimwood K, Hudson IL, Barnes GL, Bishop RF. 1992. Role of coproantibody in clinical protection of children during reinfection with rotavirus. J Clin Microbiol 30:1678-84.

166. Matson DO, O'Ryan ML, Herrera I, Pickering LK, Estes MK. 1993. Fecal antibody responses to symptomatic and asymptomatic rotavirus infections. J Infect Dis 167:577-83.

167. Franco MA, Tin C, Greenberg HB. 1997. CD8+ T cells can mediate almost complete short-term and partial long-term immunity to rotavirus in mice. J Virol 71:4165-70.

168. Franco MA, Tin C, Rott LS, VanCott JL, McGhee JR, Greenberg HB. 1997. Evidence for CD8+ T-cell immunity to murine rotavirus in the absence of perforin, fas, and gamma interferon. J Virol 71:479-86.

169. Kuklin NA, Rott L, Darling J, Campbell JJ, Franco M, Feng N, Muller W, Wagner N, Altman J, Butcher EC, Greenberg HB. 2000. alpha(4)beta(7) independent pathway for CD8(+) T cell-mediated intestinal immunity to rotavirus. J Clin Invest 106:1541-52.

170. Kuklin NA, Rott L, Feng N, Conner ME, Wagner N, Muller W, Greenberg HB. 2001. Protective intestinal anti-rotavirus B cell immunity is dependent on alpha 4 beta 7 integrin expression but does not require IgA antibody production. J Immunol 166:1894-902.

171. Youngman KR, Franco MA, Kuklin NA, Rott LS, Butcher EC, Greenberg HB. 2002. Correlation of tissue distribution, developmental phenotype, and intestinal homing receptor expression of antigen-specific B cells during the murine anti-rotavirus immune response. J Immunol 168:2173-81.

172. Mansell EA, Patton JT. 1990. Rotavirus RNA replication: VP2, but not VP6, is necessary for viral replicase activity. J Virol 64:4988-96.

173. Bican P, Cohen J, Charpilienne A, Scherrer R. 1982. Purification and characterization of bovine rotavirus cores. J Virol 43:1113-7.

174. Sandino AM, Jashes M, Faundez G, Spencer E. 1986. Role of the inner protein capsid on in vitro human rotavirus transcription. J Virol 60:797-802.

175. Helmberger-Jones M, Patton JT. 1986. Characterization of subviral particles in cells infected with simian rotavirus SA11. Virology 155:655-65.

176. Estes MKaK, A. Z. 2013. Rotaviruses, p 1374-1401. *In* Knipe DM, Howley P. M., Cohen J. I., Griffin, D. E., Lamb R. A., Martin, M. A., Racaniello, V. R., Roizman, B. (ed), Fields Virology, 6 ed, vol 1. Lippincott Williams & Wilkins, Philadelphia, PA.

177. Eichwald C, Rodriguez JF, Burrone OR. 2004. Characterization of rotavirus NSP2/NSP5 interactions and the dynamics of viroplasm formation. J Gen Virol 85:625-34.

178. Hu L, Chow DC, Patton JT, Palzkill T, Estes MK, Prasad BV. 2012. Crystallographic Analysis of Rotavirus NSP2-RNA Complex Reveals Specific Recognition of 5' GG Sequence for RTPase Activity. J Virol 86:10547-57.

179. Trask SD, Ogden KM, Patton JT. 2012. Interactions among capsid proteins orchestrate rotavirus particle functions. Curr Opin Virol 2:373-9.

180. Lopez T, Camacho M, Zayas M, Najera R, Sanchez R, Arias CF, Lopez S. 2005. Silencing the morphogenesis of rotavirus. J Virol 79:184-92.

181. Patton JT. 1996. Rotavirus VP1 alone specifically binds to the 3' end of viral mRNA, but the interaction is not sufficient to initiate minus-strand synthesis. J Virol 70:7940-7.

182. McDonald SM. 2013. RNA synthetic mechanisms employed by diverse families of RNA viruses. Wiley Interdiscip Rev RNA 4:351-67.

183. McDonald SM, Aguayo D, Gonzalez-Nilo FD, Patton JT. 2009. Shared and group-specific features of the rotavirus RNA polymerase reveal potential determinants of gene reassortment restriction. J Virol 83:6135-48.

184. Trask SD, McDonald SM, Patton JT. 2012. Structural insights into the coupling of virion assembly and rotavirus replication. Nat Rev Microbiol 10:165-77.

185. Ogden KM, Ramanathan HN, Patton JT. 2011. Residues of the rotavirus RNA-dependent RNA polymerase template entry tunnel that mediate RNA recognition and genome replication. J Virol 85:1958-69.

186. McKell AO, LaConte LE, McDonald SM. 2017. Temperature-sensitive Lesion in the N-terminal Domain of the Rotavirus Polymerase Affects its Intracellular Localization and Enzymatic Activity. J Virol doi:10.1128/JVI.00062-17.

187. Busi C, Martella V, Papetti A, Sabelli C, Lelli D, Alborali GL, Gibelli L, Gelmetti D, Lavazza A, Cordioli P, Boniotti MB. 2017. Group A Rotavirus Associated with Encephalitis in Red Fox. Emerg Infect Dis 23:1535-1538.

188. Greenberg HB, Vo PT, Jones R. 1986. Cultivation and characterization of three strains of murine rotavirus. J Virol 57:585-90.

189. He B, Huang X, Zhang F, Tan W, Matthijnssens J, Qin S, Xu L, Zhao Z, Yang L, Wang Q, Hu T, Bao X, Wu J, Tu C. 2017. Group A Rotaviruses in Chinese Bats: Genetic Composition, Serology, and Evidence for Bat-to-Human Transmission and Reassortment. J Virol 91.

190. Yinda CK, Zeller M, Conceicao-Neto N, Maes P, Deboutte W, Beller L, Heylen E, Ghogomu SM, Van Ranst M, Matthijnssens J. 2016. Novel highly divergent reassortant bat rotaviruses in Cameroon, without evidence of zoonosis. Sci Rep 6:34209.

191. Esona MD, Roy S, Rungsrisuriyachai K, Gautam R, Hermelijn S, Rey-Benito G, Bowen MD. 2018. Molecular characterization of a human G20P[28] rotavirus a strain with multiple genes related to bat rotaviruses. Infect Genet Evol 57:166-170.

192. Matthijnssens J, Van Ranst M. 2012. Genotype constellation and evolution of group A rotaviruses infecting humans. Curr Opin Virol 2:426-33.

193. Heiman EM, McDonald SM, Barro M, Taraporewala ZF, Bar-Magen T, Patton JT. 2008. Group A human rotavirus genomics: evidence that gene constellations are influenced by viral protein interactions. J Virol 82:11106-16.

194. Ghosh S, Gatheru Z, Nyangao J, Adachi N, Urushibara N, Kobayashi N. 2011. Full genomic analysis of a simian SA11-like G3P[2] rotavirus strain isolated from an asymptomatic infant: identification of novel VP1, VP6 and NSP4 genotypes. Infect Genet Evol 11:57-63.

195. De Grazia S, Giammanco GM, Doro R, Bonura F, Marton S, Cascio A, Martella V, Banyai K. 2015. Identification of a multi-reassortant G12P[9] rotavirus with novel VP1, VP2, VP3 and NSP2 genotypes in a child with acute gastroenteritis. Infect Genet Evol 35:34-7.

196. Pedregosa FV, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, E. 2011. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research 12:2825-3830.

197. Pagel M. 1999. The Maximum Likelihood Approach to Reconstructing Ancestral Character States of Discrete Characters on Phylogenies. Systematic Biology 48:612-622.

198. Zeng CQ, Estes MK, Charpilienne A, Cohen J. 1998. The N terminus of rotavirus VP2 is necessary for encapsidation of VP1 and VP3. J Virol 72:201-8.

199. McDonald SM, Patton JT. 2008. Molecular characterization of a subgroup specificity associated with the rotavirus inner capsid protein VP2. J Virol 82:2752-64.

200. Zeng CQ, Labbe M, Cohen J, Prasad BV, Chen D, Ramig RF, Estes MK. 1994. Characterization of rotavirus VP2 particles. Virology 201:55-65.

201. Boudreaux CE, Vile DC, Gilmore BL, Tanner JR, Kelly DF, McDonald SM. 2013. Rotavirus core shell subdomains involved in polymerase encapsidation into virus-like particles. J Gen Virol 94:1818-26.

202. Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30:772-80.

203. Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312-3.

204. Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, Lindahl E. 2015. GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX 1:19-25.

205. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. 2004. UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem 25:1605-12.

206. Mingo R, Zhang S, Long CP, LaConte LEW, McDonald SM. 2017. Genetic determinants restricting the reassortment of heterologous NSP2 genes into the simian rotavirus SA11 genome. Sci Rep 7:9301.

207. McDonald SM, Tao YJ, Patton JT. 2009. The ins and outs of four-tunneled Reoviridae RNA-dependent RNA polymerases. Curr Opin Struct Biol 19:775-82.

208. Ogden KM, Ramanathan HN, Patton JT. 2012. Mutational analysis of residues involved in nucleotide and divalent cation stabilization in the rotavirus RNA-dependent RNA polymerase catalytic pocket. Virology 431:12-20.

209. Kanai Y, Komoto S, Kawagishi T, Nouda R, Nagasawa N, Onishi M, Matsuura Y, Taniguchi K, Kobayashi T. 2017. Entirely plasmid-based reverse genetics system for rotaviruses. Proc Natl Acad Sci U S A doi:10.1073/pnas.1618424114.

210. Pelletier H, Sawaya MR, Kumar A, Wilson SH, Kraut J. 1994. Structures of ternary complexes of rat DNA polymerase beta, a DNA template-primer, and ddCTP. Science 264:1891-903.

211. Doublie S, Sawaya MR, Ellenberger T. 1999. An open and closed case for all polymerases. Structure 7:R31-5.

212. Steitz TA. 1998. A mechanism for all polymerases. Nature 391:231-2.

213. Westover KD, Bushnell DA, Kornberg RD. 2004. Structural basis of transcription: nucleotide selection by rotation in the RNA polymerase II active center. Cell 119:481-9.

214. Yin YW, Steitz TA. 2004. The structural mechanism of translocation and helicase activity in T7 RNA polymerase. Cell 116:393-404.

215. Ng KK, Arnold JJ, Cameron CE. 2008. Structure-function relationships among RNA-dependent RNA polymerases. Curr Top Microbiol Immunol 320:137-56.

216. Butcher SJ, Grimes JM, Makeyev EV, Bamford DH, Stuart DI. 2001. A mechanism for initiating RNA-dependent RNA polymerization. Nature 410:235-40.

217. Liang B, Li Z, Jenni S, Rahmeh AA, Morin BM, Grant T, Grigorieff N, Harrison SC, Whelan SPJ. 2015. Structure of the L Protein of Vesicular Stomatitis Virus from Electron Cryomicroscopy. Cell 162:314-327.

218. Reich S, Guilligay D, Pflug A, Malet H, Berger I, Crepin T, Hart D, Lunardi T, Nanao M, Ruigrok RW, Cusack S. 2014. Structural insight into cap-snatching and RNA synthesis by influenza polymerase. Nature 516:361-6.

219. Ng KK, Pendas-Franco N, Rojo J, Boga JA, Machin A, Alonso JM, Parra F. 2004. Crystal structure of norwalk virus polymerase reveals the carboxyl terminus in the active site cleft. J Biol Chem 279:16638-45.

220. Zhao B, Yi G, Du F, Chuang YC, Vaughan RC, Sankaran B, Kao CC, Li P. 2017. Structure and function of the Zika virus full-length NS5 protein. Nat Commun 8:14762.

221. Venkataraman S, Prasad B, Selvarajan R. 2018. RNA Dependent RNA Polymerases: Insights from Structure, Function and Evolution. Viruses 10.

222. O'Reilly EK, Kao CC. 1998. Analysis of RNA-dependent RNA polymerase structure and function as guided by known polymerase structures and computer predictions of secondary structure. Virology 252:287-303.

223. Ahlquist P. 2006. Parallels among positive-strand RNA viruses, reverse-transcribing viruses and double-stranded RNA viruses. Nat Rev Microbiol 4:371-82.

224. Beck J, Nassal M. 2007. Hepatitis B virus replication. World J Gastroenterol 13:48-64.

225. Thompson AA, Peersen OB. 2004. Structural basis for proteolysis-dependent activation of the poliovirus RNA-dependent RNA polymerase. EMBO J 23:3462-71.

226. Hansen JL, Long AM, Schultz SC. 1997. Structure of the RNA-dependent RNA polymerase of poliovirus. Structure 5:1109-22.

227. Lyle JM, Bullitt E, Bienz K, Kirkegaard K. 2002. Visualization and functional analysis of RNA-dependent RNA polymerase lattices. Science 296:2218-22.

228. Pflug A, Lukarska M, Resa-Infante P, Reich S, Cusack S. 2017. Structural insights into RNA synthesis by the influenza virus transcription-replication machine. Virus Res 234:103-117.

229. Green TJ, Luo M. 2009. Structure of the vesicular stomatitis virus nucleocapsid in complex with the nucleocapsid-binding domain of the small polymerase cofactor, P. Proc Natl Acad Sci U S A 106:11713-8.

230. Lu S, Li S, Zhang J. 2014. Harnessing allostery: a novel approach to drug discovery. Med Res Rev 34:1242-85.

231. Das K, Arnold E. 2013. HIV-1 reverse transcriptase and antiviral drug resistance. Part 1. Curr Opin Virol 3:111-8.

232. Campagnola G, Gong P, Peersen OB. 2011. High-throughput screening identification of poliovirus RNA-dependent RNA polymerase inhibitors. Antiviral Res 91:241-51.

233. Lim SP, Noble CG, Seh CC, Soh TS, El Sahili A, Chan GK, Lescar J, Arora R, Benson T, Nilar S, Manjunatha U, Wan KF, Dong H, Xie X, Shi PY, Yokokawa F. 2016. Potent Allosteric Dengue Virus NS5 Polymerase Inhibitors: Mechanism of Action and Resistance Profiling. PLoS Pathog 12:e1005737.

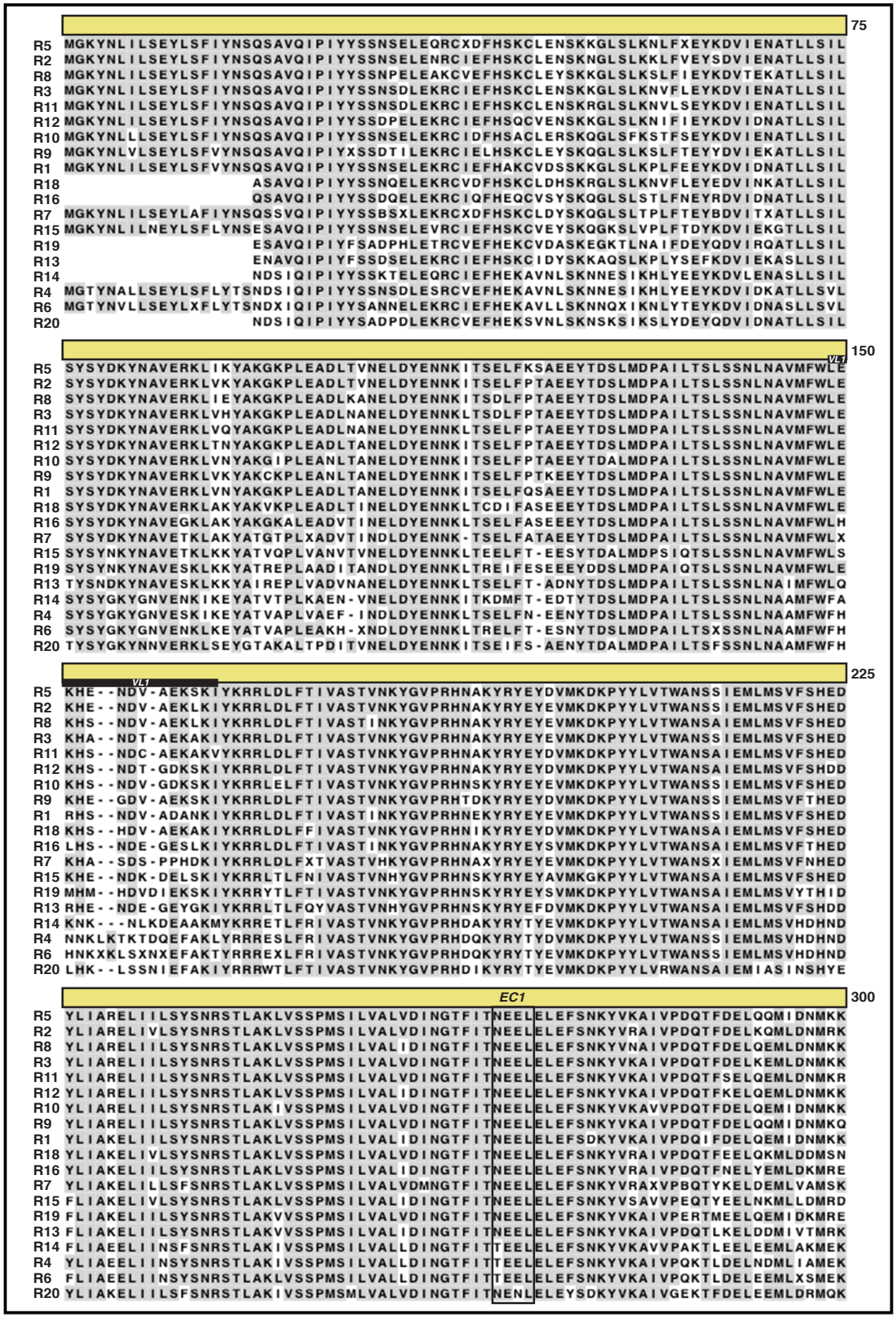
234. Deval J, Jin Z, Chuang YC, Kao CC. 2017. Structure(s), function(s), and inhibition of the RNA-dependent RNA polymerase of noroviruses. Virus Res 234:21-33.

Appendix A. Table of GenBank Accession Numbers for VP1 and VP2 Sequences.

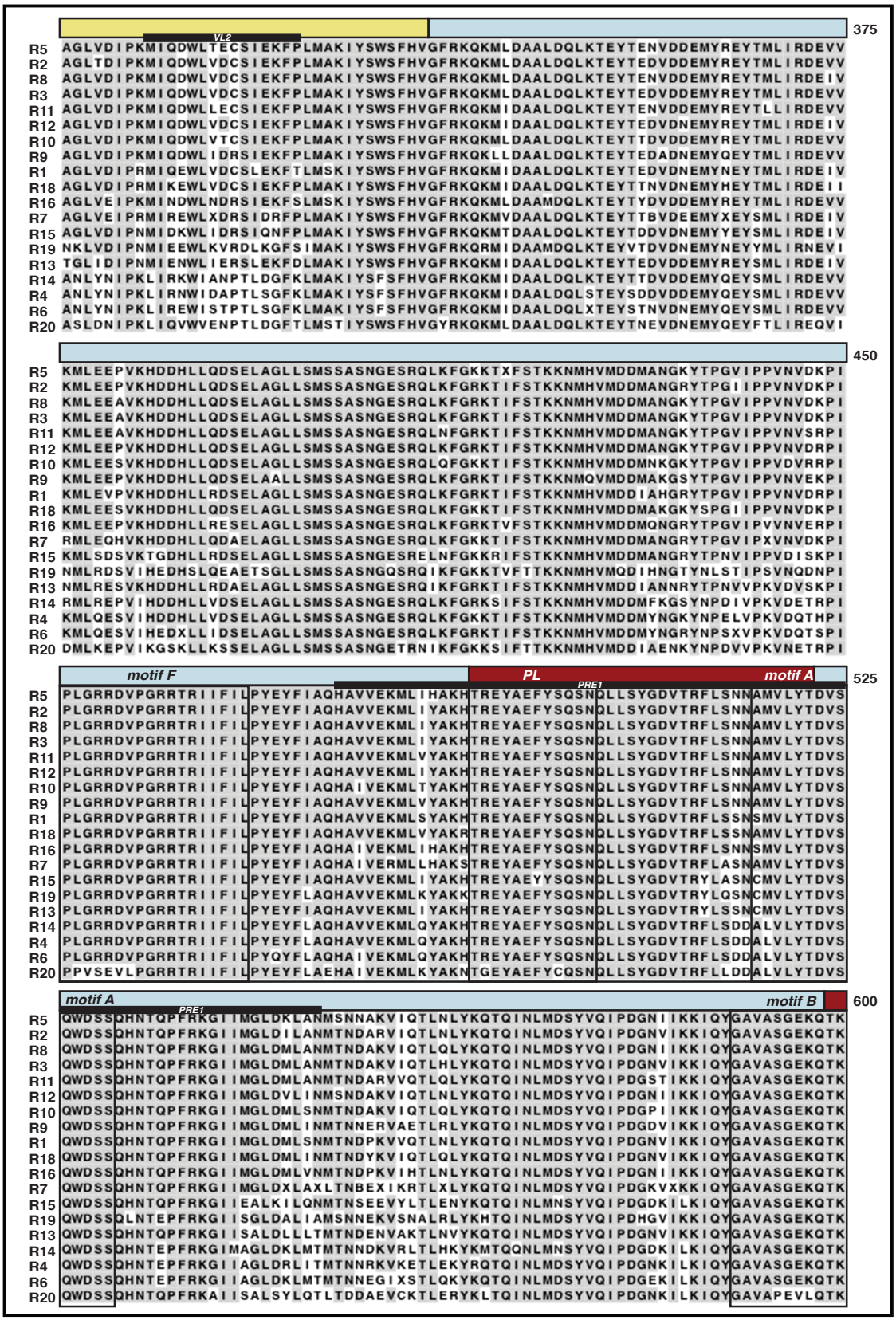
Full Strain Name	VP1 accession #		VP2 accession #	
RVA/Human-wt/BEL/B4633/2003/G12P[8]	DQ146638	R1	DQ146639	C1
RVA/Human-tc/JPN/YO/1977/G3P[8]	DQ870497	R1	DQ870498	C1
RVA/Human-wt/BEL/B3458/2003/G9P[8]	DQ870501	R1	DQ870502	C1
RVA/Human-wt/DEU/GER126-08/2008/G12P[8]	FJ747613	R1	FJ747614	C1
RVA/Human-wt/DEU/GER172-08/2008/G12P[6]	FJ747625	R1	FJ747626	C1
RVA/Pig-wt/IND/RU172/2002/G12P[7]	GU199191	R1	GU199192	C1
RVA/Human-wt/NPL/KTM368/2004/G11P[25]	GU199492	R1	GU199493	C1
RVA/Human-wt/BGD/Matlab36/2002/G11P[8]	GU199503	R1	GU199504	C1
RVA/Human-wt/BEL/BE00030/2008/G1P[8]	HQ392245	R1	HQ392254	C1
RVA/Panda-tc/CHN/CH-1/2009/G1P[7]	HQ641297	R1	HQ641294	C1
RVA/Pig-tc/KOR/PRG9235/2006/G9P[23]	JF796701	R1	JF796702	C1
RVA/Pig-tc/KOR/PRG921/2006/G9P[23]	JF796712	R1	JF796713	C1
RVA/Pig-tc/KOR/PRG942/2006/G9P[23]	JF796723	R1	JF796724	C1
RVA/Pig-tc/KOR/PRG9121/2006/G9P[23]	JF796734	R1	JF796735	C1
RVA/Human-tc/THA/Mc323/1989/G9P[19]	JN104619	R1	JN104620	C1
RVA/Human-wt/USA/2008747288/2008/G1P[8]	JN258375	R1	JN258385	C1
RVA/Human-wt/CMR/MRC-DPRU1424/2009/G9P[8	JN605404	R1	JN605405	C1
RVA/Human-wt/THA/CU460-KK/2009/G12P[8]	JN706453	R1	JN706468	C1
RVA/Human-wt/THA/CU938-BK/2011/G3P[8]	JN706459	R1	JN706474	C1
RVA/Human-tc/CHN/Y128/2004/G1P[8]	JQ087423	R1	JQ087424	C1
RVA/Human-wt/Croatia/CR2006/2006/G8P[8]	JQ988899	R1	JQ988900	C1
RVA/Human-wt/BEL/BE2001/2009/G9P[6]	JQ993321	R1	JQ993322	C1
RVA/Human-wt/AUS/CK00100/2010/G1P[8]	JX027974	R1	JX027977	C1
RVA/Human-tc/USA/Wa/1974/G1P[8]	JX406747	R1	JX406748	C1
RVA/Pig-wt/ITA/2CR/2009/G9P[23]	KC610678	R1	KC610679	C1
RVA/Human-wt/BRB/CDC1133/2012/G4P[14]	KF035107	R1	KF035108	C1
RVA/Human-wt/CMR/MRC-DPRU1417/2009/G9P[8	KF636289	R1	KF636290	C1
RVA/Human-wt/HUN/BP271/2000/G4P[6]	KF835889	R1	KF835897	C1
RVA/Human-wt/IND/B08299/2007/G11P[25]	KJ501692	R1	KJ501693	C1
RVA/Human-wt/PRY/954SR/2005/G4P[8]	KJ559301	R1	KJ559302	C1
RVA/Human-wt/BFA/MRC-DPRU2862/2009/G1P[8]	KJ751823	R1	KJ751824	C1
RVA/Human-wt/GMB/MRC-DPRU3174/2010/G1P[8	KJ752284	R1	KJ752285	C1
RVA/Human-wt/COD/KisB332/2008/G4P[6]	KJ870900	R1	KJ870901	C1
RVA/Pig-tc/BEL/RV277/1977/G1P[7]	KM820701	R1	KM820708	C1
RVA/Human-wt/ETH/MRC-DPRU4970/2010/G12P[8	KP752790	R1	KP752791	C1
RVA/Pig-wt/UGA/KYE-14-A047/2014/G3P[13]	KX988264	R1	KX988265	C1
RVA/Pig-wt/UGA/KYE-14-A048/2014/G3P[13]	KX988275	R1	KX988276	C1
RVA/Simian-wt/KNA/08979/2015/G5P[X]	KY053144	R1	KY053145	C1
RVA/Pig-wt/UGA/BUW-14-A008/2014/G12P[8]	KY055416	R1	KY055417	C1
RVA/Rabbit-tc/NLD/K1130027/2011/G6P[11]	KC488885	R2	KC488886	C2
RVA/Camel-wt/SDN/MRC-DPRU447/2002/G8P[11]	KC257091	R2	KC257092	C2
RVA/Human-wt/GHA/GH018-08/2008/G8P[6]	KJ748465	R2	KJ748466	C2
RVA/Human-wt/PRY/492SR/2004/G8P[1]	KJ803835	R2	KJ803836	C2
RVA/Human-wt/US/09US7118/2009/G3P[24]	KF541281	R2	KF541282	C3
RVA/Human-wt/BRA/MS11142/2005/G2P[4]	KJ721726	R2	KJ940055	C2
RVA/Cow-wt/ZAF/MRC-DPRU456/2009/G6P[11]	KP752877	R2	KP752878	C2
RVA/Pig-tc/THA/P343/1991/G10P[5]	AB972856	R2	AB972857	C2
RVA/Human-wt/THA/PCB-180/2013/G1P[8]	LC066639	R2	LC066640	C2
RVA/Human-wt/THA/LS-04/2013/G2P[8]	LC086736	R2	LC086737	C2
RVA/Cow-tc/THA/A5-13/1988/G8P[1]	LC133525	R2	LC133526	C2
RVA/Bat-wt/ZMB/LUS12-14/2012/G3P[3]	LC158119	R2	LC158120	C2
RVA/Human-wt/TUN/17237/2008/G6P[9]	JX271001	R2	JX271002	C2
RVA/Human-wt/BEL/BE5028/2012/G3P[14]	KP258398	R2	KP258399	C2
RVA/Human-wt/USA/2012741499/2012/G24P[14]	KT281120	R2	KT281121	C2
RVA/Rabbit-tc/ITA/30-96/1996/G3P[14]	DQ205221	R2	DQ205222	C2
RVA/Human-tc/USA/DS-1/1976/G2P[4]	HQ650116	R2	HQ650117	C2
RVA/Cow-tc/USA/NCDV/1967/G6P6[1]	DQ870493	R2	DQ870494	C2
RVA/Human-tc/AUS/MG6/1993/G6P[14]	EF554093	R2	EF554094	C2
RVA/Human-wt/HUN/Hun5/1997/G6P[14]	EF554104	R2	EF554105	C2
RVA/Human-tc/ITA/PA169/1988/G6P[14]	EF554126	R2	EF554127	C2
RVA/Sheep-tc/ESP/OVR762/2002/G8P[14]	EF554148	R2	EF554149	C2
RVA/Simian-tc/USA/RRV/1975/G3P[3]	EF583006	R2	EF583007	C3
RVA/Antelope-wt/ZAF/RC-18-08/G6P[14]	FJ495126	R2	FJ495127	C2
RVA/Cat-wt/ITA/BA222/2005/G3P[9]	GU827406	R2	GU827407	C2
RVA/Horse-wt/ARG/E30/1993/G3P[12]	JF712566	R2	JF712567	C2
RVA/Horse-wt/ARG/E4040/2008/G14P[12]	JN872865	R2	JN872866	C2
RVA/Horse-wt/IRL/04V2024/2004/G14P[12]	JN903527	R2	JN903525	C2
RVA/Simian-tc/ZAF/SA11-N2/1958/G3P[2]	JN827244	R2	JN827245	C5
RVA/Simian-tc/ZAF/SA11-N5/1958/G3P[2]	JQ688673	R2	JQ688674	C5
RVA/Human-wt/USA/12US1134/2012/G3P[9]	KF500519	R2	KF500520	C2
RVA/Human-wt/GTM/2009726790/2009/G8P[14]	KP006506	R2	KP006507	C2
RVA/Cow-tc/VEN/BRV033/1990/G6P[1]	EF560612	R2	EF560613	C2
RVA/Human-wt/IND/N155/2003/G10P[11]	EU200793	R2	EU200794	C2
RVA/Macaque-tc/USA/PTRV/1990/G8P[1]	FJ422131	R2	FJ422132	C2
RVA/Sheep-tc/CHN/Lamb-NT/2007/G10P[15]	FJ031024	R2	FJ031025	C2
RVA/Giraffe-wt/IRL/GirRV/2007/G10P[11]	GQ428141	R2	GQ428142	C2
RVA/Human-tc/KEN/B12/1987/G8P[1]	HM627542	R2	HM627543	C2
RVA/Horse/IRL/FI-14/2003/G3P[12]	JN903528	R2	JN903526	C2
RVA/Goat-wt/CHN/XL/2010/G10P[15]	JQ004970	R2	JQ004971	C2

Full Strain Name	VP1 accession #		VP2 accession #	
RVA/Human-wt/USA/LB1562/2010/G9P[4]	KC782519	R2	KC782520	C2
RVA/Horse-tc/JPN/OH-4/1982/G6P[5]	KC815658	R2	KC815659	C2
RVA/Horse-tc/JPN/HH-22/1989/G3P[12]	KC815669	R2	KC815670	C2
RVA/Horse-tc/JPN/BI/1981/G3P[12]	KC815680	R2	KC815681	C2
RVA/Human-tc/KOR/CAU12-2-51/2012/G3P[9]	KJ187599	R2	KJ187600	C2
RVA/Human-wt/CMR/MA109/2011/2010/G3P[6]	KM660256	R2	KM660287	C2
RVA/Deer-wt/SLO/D38-14/2014/G6P[15]	KU708259	R2	KU708260	C2
RVA/Deer-wt/SLO/D110-15/2015/G8P[14]	KY426809	R2	KY426810	C2
RVA/Cow-tc/JPN/Dai-10/2007/G24P[33]	AB573070	R2	AB573071	C2
RVA/Human-wt/VNM/1149/2014/G8P[8]	LC074734	R2	LC074735	C2
RVA/Horse-wt/ARG/E3198/2008/G3P[3]	JX036365	R3	JX036366	C3
RVA/Human-tc/CHN/L621/2006/G3P[9]	JX946159	R3	JX946160	C3
RVA/Human-wt/CHN/E2451/2011/G3P[9]	JX946168	R3	JX946169	C3
RVA/Human-wt/AUS/RCH272/2012/G3P[14]	KF690125	R3	KF690126	C3
RVA/Human-wt/PRY/1471SR/2006/G12P[9]	KJ412575	R3	KJ412576	C3
RVA/Human-wt/PRY/1701SR/2009/G1P[9]	KJ412620	R3	KJ412621	C3
RVA/Human-wt/PRY/1093A/2007/G12P[4]	KJ412721	R3	KJ412722	C3
RVA/Human-wt/PRY/985A/2006/G12P[8]	KJ412808	R3	KJ412809	C3
RVA/Human-wt/PRY/1259A/2009/G12P[9]	KJ412819	R3	KJ412820	C3
RVA/Human-wt/US/2012841174/2012/G8P[14]	KJ411432	R3	KJ411433	C2
RVA/Human-wt/BRA/PE18974/2010/G12P[9]	KF907277	R3	KF907284	C3
RVA/Human-wt/BRA/PE15776/2008/G12P[9]	KF907278	R3	KF907283	C3
RVA/Human-tc/BRA/R55/1997/G3P[9]	KJ820847	R3	KJ820848	C3
RVA/Human-tc/BRA/R57/1997/G3P[9]	KJ820858	R3	KJ820859	C3
RVA/Cat-tc/JPN/FRV317/1994/G3P[9]	LC328226	R3	LC328232	C3
RVA/Cat-tc/JPN/FRV303/1993/G3P[3]	LC328222	R3	LC328228	C2
RVA/Human-tc/THA/T152/1998/G12P[9]	DQ146699	R3	DQ146700	C3
RVA/Human-tc/JPN/AU-1/1982/G3P3[9]	DQ490533	R3	DQ490536	C3
RVA/Rhesus-tc/USA/TUCH/2002/G3P[24]	EF583010	R3	EF583011	C3
RVA/Human-tc/ITA/PA260-97/1997/G3P[3]	HQ661112	R3	HQ661113	C3
RVA/Dog-tc/ITA/RV52-96/1996/G3P[3]	HQ661123	R3	HQ661124	C2
RVA/Dog-tc/ITA/RV198-95/1995/G3P[3]	HQ661134	R3	HQ661135	C2
RVA/Human-tc/JPN/K8/1977/G1P[9]	JQ713645	R3	JQ713646	C3
RVA/Bat/MSLH14/G3P[3]	KC960619	R3	KC960620	C3
RVA/Bat-wt/CHN/LZHP2/2015/G3P[3]	KX814945	R3	KX814946	C3
RVA/Human-tc/Ro1845/ISR/1985/G3P[3]	EU708890	R3	EU708891	C2
RVA/Human-tc/USA/HCR3A/1984/G3P[3]	EU708901	R3	EU708902	C2
RVA/Dog-tc/USA/CU-1/1980/G3P[3]	EU708912	R3	EU708913	C2
RVA/Dog-tc/USA/K9/1979/G3P[3]	EU708923	R3	EU708924	C2
RVA/Dog-tc/USA/A79-10/1979/G3P[3]	EU708934	R3	EU708935	C2
RVA/Human-wt/XXX/CU365-KK/2008/G3P[9]	JN706445	R3	JN706480	C3
RVA/Rabbit-tc/CHN/N5/1992/G3P[14]	JQ423902	R3	JQ423903	C3
RVA/Bat-tc/CHN/MYAS33/2013/G3P[10]	KJ020891	R3	KJ020892	C3
RVA/Human-wt/KOR/CAU14-1-262/2014/G3P[9]	KR262149	R3	KR262150	C3
RVA/Bat-wt/CHN/BSTM70/2015/G3P[3]	KX814924	R3	KX814925	C3
RVA/Chicken-tc/xxx/BRS/115/xxx/G7P[35]	KJ725017	R4	KJ725018	C4
RVA/VelvetScoter-tc/JPN/RK1/1989/G18P[17]	LC088096	R4	LC088097	C4
RVA/Turkey-tc/IRL/Ty-1/1979/G17P[38]	LC088107	R4	LC088108	C4
RVA/Turkey-tc/IRL/Ty-3/1979/G7P[35]	LC088115	R4	LC088116	C4
RVA/Turkey-tc/GER/03V0002E10/2003/G22P[35]	JX204822	R4	JX204823	C4
RVA/Pheasant-tc/GER/10V0112H5/2010/G23P[37]	JX204811	R4	JX204812	C4
RVA/Fox-wt/ITA/288356/2011/G18P[17]	KT873808	R4	KT873809	C4
RVA/Pheasant-wt/HUN/216/2015/G23P[37]	KU587853	R4	KU587854	C4
RVA/Pigeon-tc/JPN/PO-13/1983/G18P[17]	AB009629	R4	AB009630	C4
RVA/Goat-wt/ARG/0040/2011/G8P[1]	KF577838	R5	KF577839	C2
RVA/Alpaca-wt/PER/356/2010/G3P[14]	KT878995	R5	KT878996	Cx
RVA/Alpaca-wt/PER/562/2010/G8P[14]	KT887520	R5	KT887521	Cx
RVA/Guanaco-wt/ARG/Chubut/1999/G8P[14]	FJ347100	R5	FJ347101	C2
RVA/Cow-wt/ARG/B383/1998/G15P[11]	FJ347111	R5	FJ347112	C2
RVA/Guanaco-wt/ARG/Rio_Negro/1998/G8P[1]	FJ347122	R5	FJ347123	C2
RVA/Chicken-tc/DEU/02V0002G3/2002/G19P[30]	FJ169853	R6	FJ169854	C6
RVA/Chicken/KOR/D62/2013/G19P[31]	KM254178	R6	KM254179	C6
RVAMouse/Stanford/EB/1995/G16P[16]	HQ540507	R7	HQ540508	C7
RVA/Mouse-tc/USA/ETD_822/XXXX/G16P[16]	HQ540509	R7	HQ540510	C7
RVA/Human-tc/KEN/B10/1987/G3P[2]	HM627553	R8	HM627554	C5
RVA/Horse-tc/L338/GBR/1991/GxP[18]	JF712555	R9	JF712556	C9
RVA/SugarGlider-tc/JPN/SG385/2012/G27P[36]	AB971760	R10	AB971761	C10
RVA/Rat-wt/XXX/KS/11/0573/G3P[3]	KJ879448	R11	KJ879449	C11
RVA/Human-xx/ITA/ME848/2012/G12P[9]	KR632623	R12	KR632624	C12
RVA/Human-wt/SUR/2014735512/2013/G20P[28]	KX257410	R13	KX257409	C13
RVA/CommonGull-wt/JPN/Ho374/2013/G28P[39]	LC088218	R14	LC088219	C14
RVA/Bat-wt/CMR/BatLi10/2014/G30P[42]	KX268743	R15	KX268744	C15
RVA/Bat-wt/CMR/BatLi09/2014/G30P[42]	KX268754	R15	KX268755	C15
RVA/Bat-wt/CMR/BatLi08/2014/G31P[42]	KX268765	R15	KX268766	C15
RVA/Bat-wt/CMR/BatLy17/2014/G31P[47]	KX268787	R15	KX268788	C15
RVA/Bat-wt/CMR/BatLy03/2014/G25P[42]	KX268776	R16	KX268777	C8
RVA/Shrew-wt/CHN/LW9/2013/G32P[46]	KU243542	R18	KU243571	C17
RVA/Bat-wt/CHN/GLRL1/2005/G33P[48]	KX814935	R19	KX814936	C18
RVA/Bat-wt/CHN/YSSK5/2015/G3P[3]	KX814956	R20	KX814957	C2

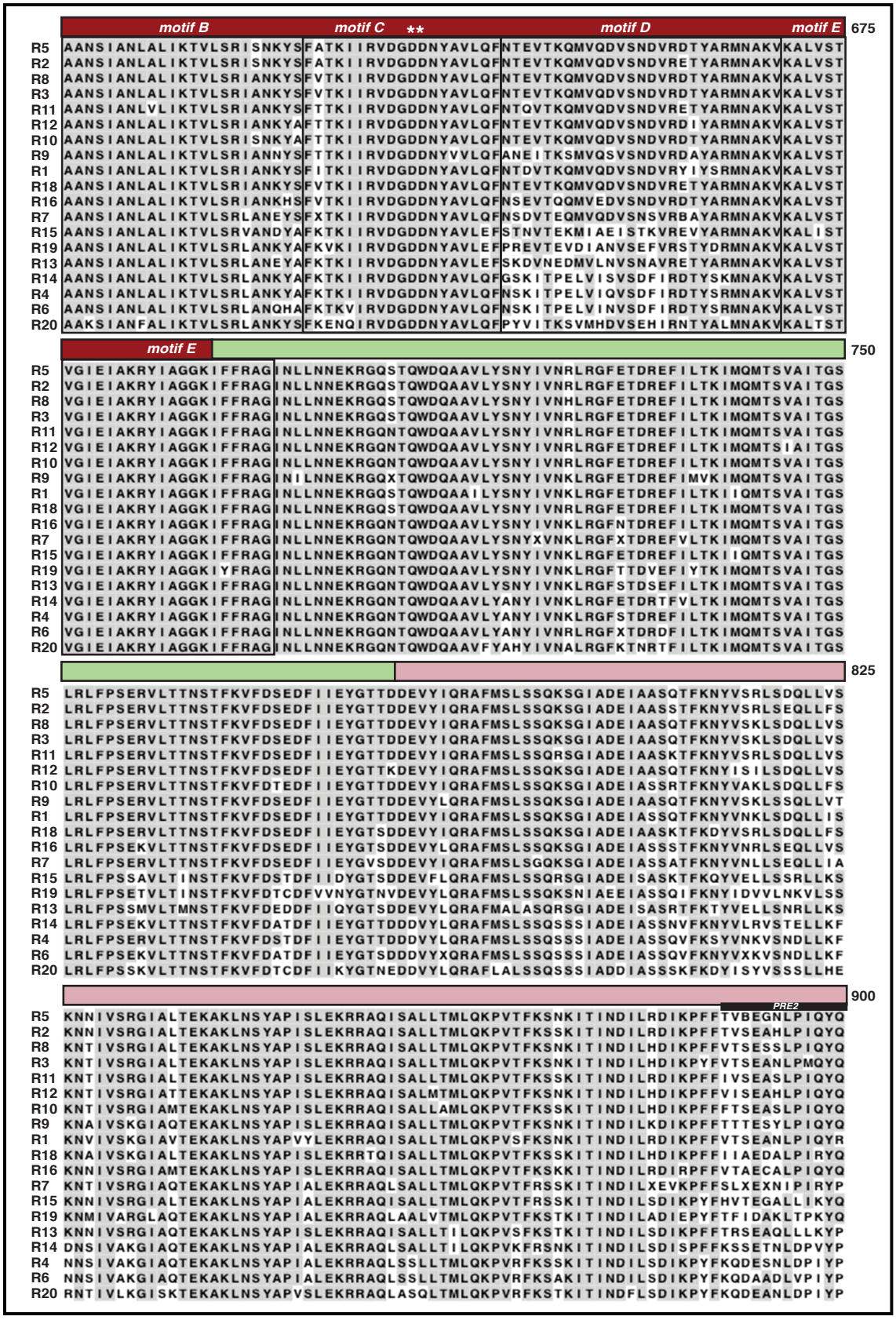
Appendix B. VP1 genotype consensus amino acid sequence alignment. Page 1 of 4

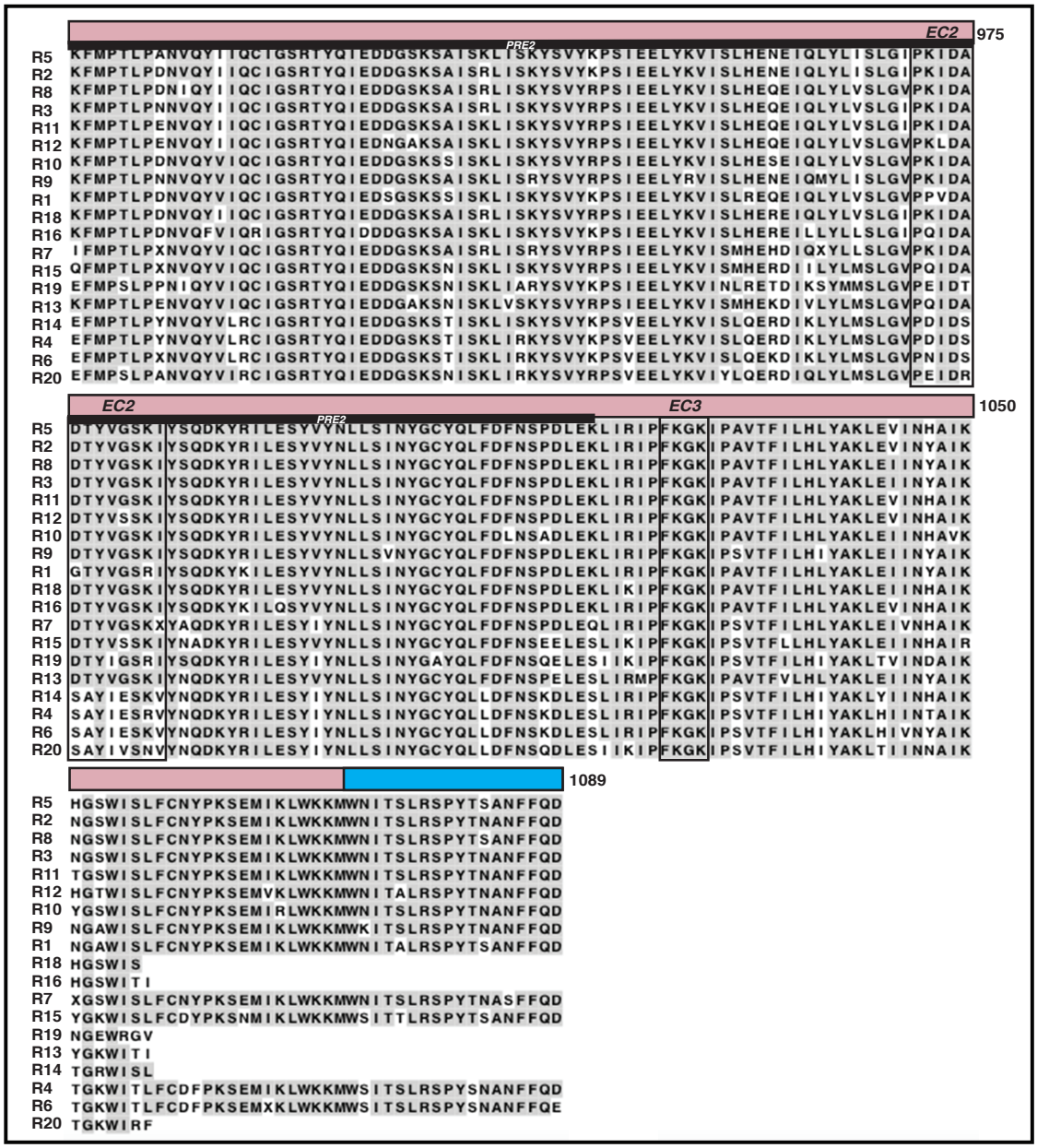


Appendix B. VP1 genotype consensus amino acid sequence alignment. Page 2 of 4



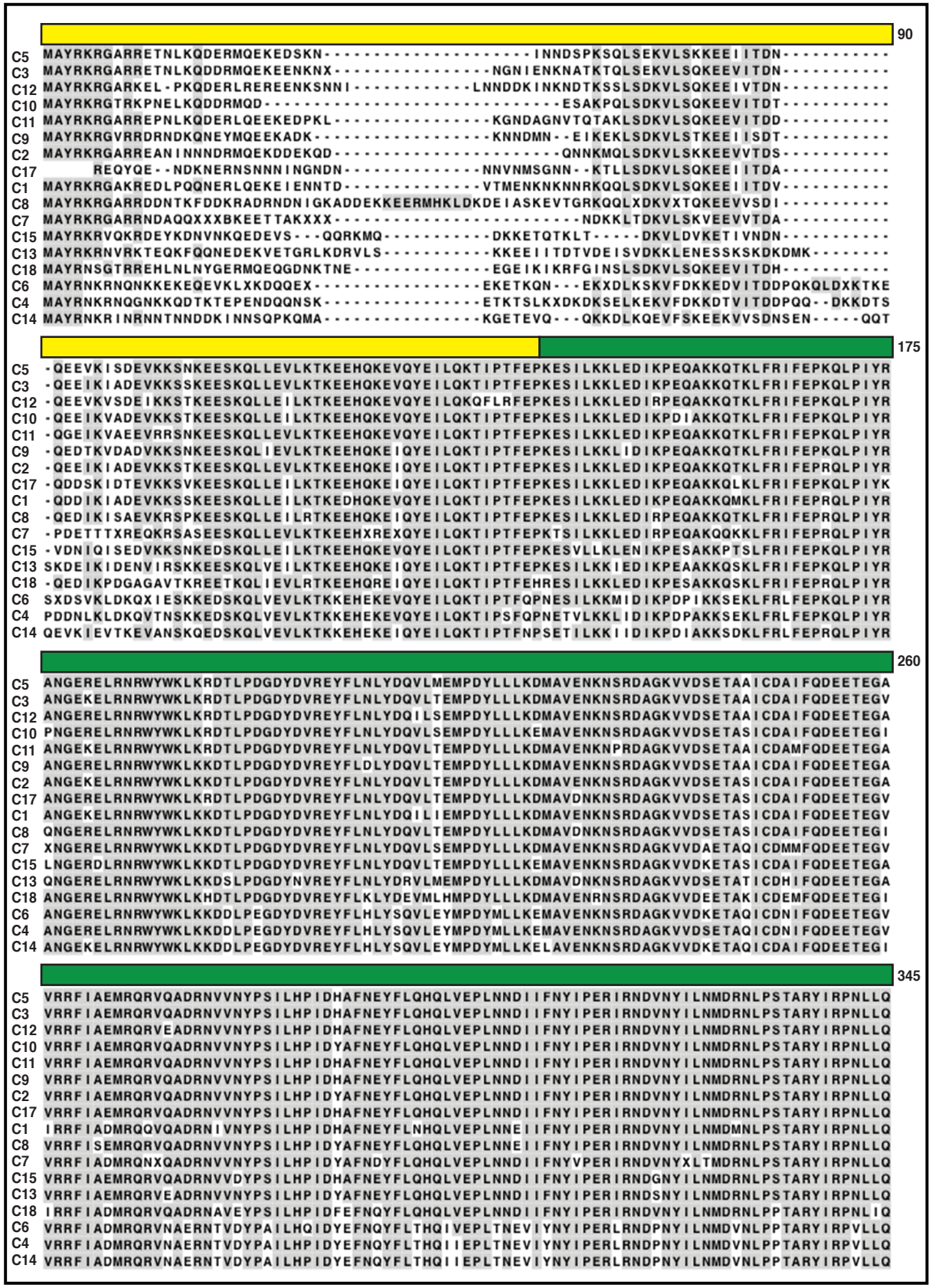
Appendix B. VP1 genotype consensus amino acid sequence alignment. Page 3 of 4



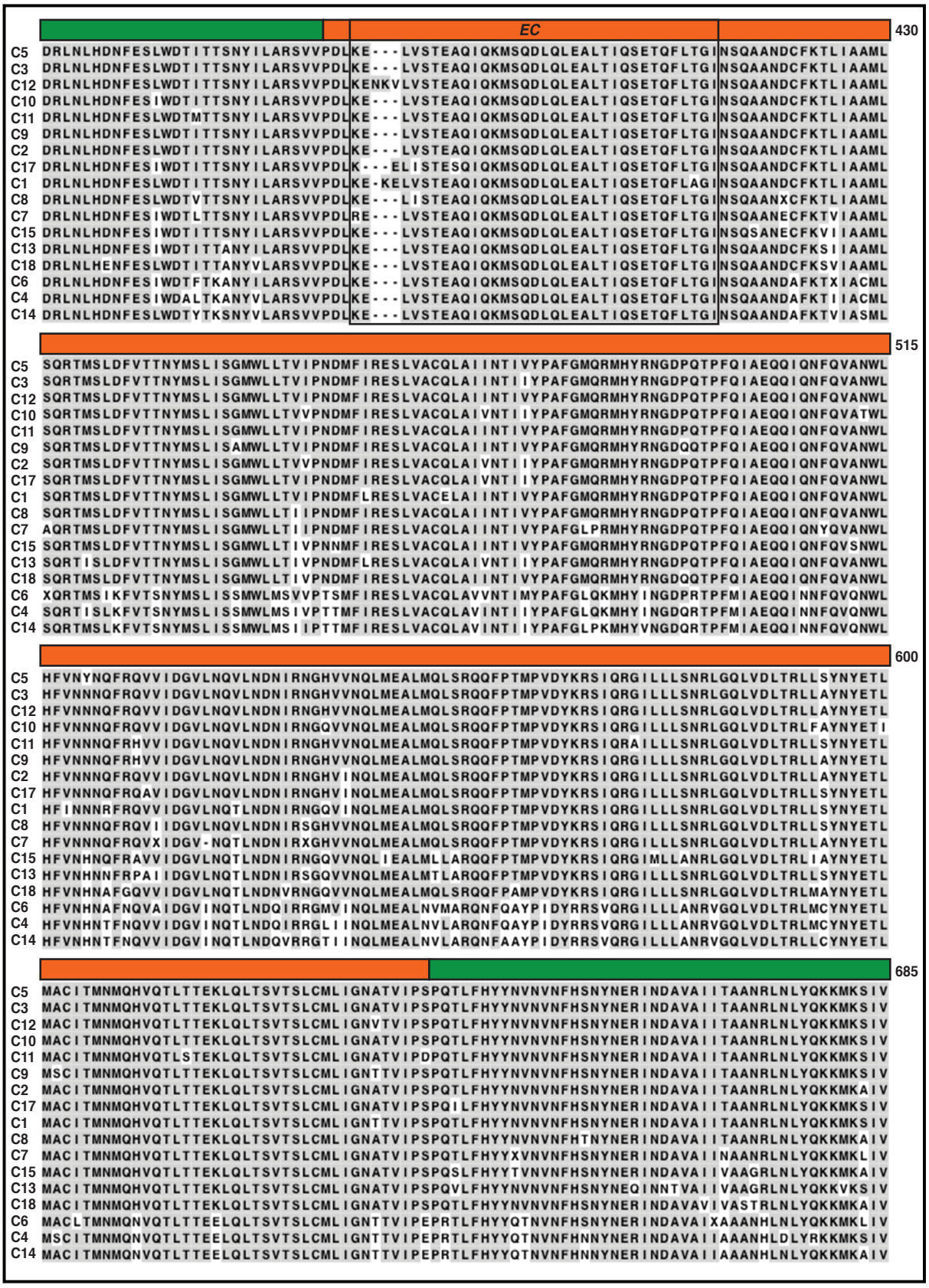


Appendix B. VP1 genotype consensus amino acid sequence alignment. The schematic shows an amino acid sequence alignment of RVA VP1 consensus sequences for genotypes R1-R16 and R18-R20. The VP1 domains and subdomains are represented by a line above the sequence and are colored as in Fig. 3-1A-B. Motifs A-F, the priming loop (PL), and the VP1-VP2 interaction sites predicted by Estrozi et al. (EC1, EC2, and EC3) are outlined in boxes. Surface-exposed regions surrounding the +RNA exit tunnel (PRE) and two surface-exposed loops (VL) are indivated in thick blacklines above the sequence. The active site is indicated with asterisks. Amino acid positions are indicated to the right. Dashes indicate gaps in the protein sequence, light gray shading indicates conservation of amino acid identity, and white shading represents variation in amino acid identity. Genotypes are listed on the left. R20 (lineage I) is positioned at the bottom of the alignment due to its amino acid similarities with lineage II genotypes (R4, R6, and R14).

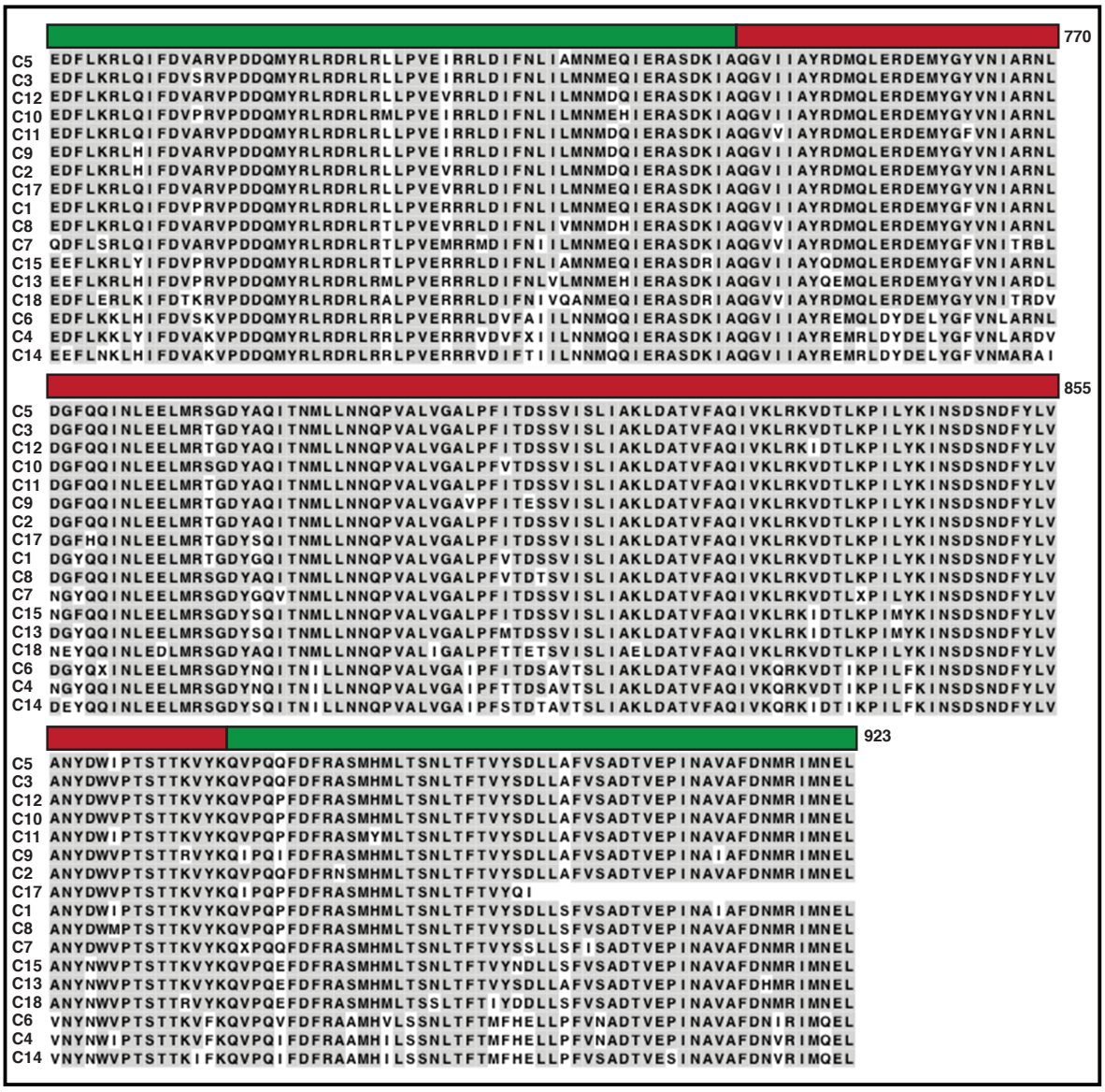
Appendix C. VP2 genotype consensus amino acid sequence alignment. Page 1 of 3



Appendix C. VP2 genotype consensus amino acid sequence alignment. Page 2 of 3



Appendix C. VP2 genotype consensus amino acid sequence alignment. Page 3 of 3



Appendix C. VP2 genotype consensus amino acid sequence alignment. The schematic shows an amino acid sequence alignment of RVA VP2 consensus sequences for genotypes C1-C15 and C17-C18. Dashes indicate gaps in the protein sequence, light gray shading indicates conservation of amino acid identity, and white shading represents variation in amino acid identity. Genotypes are listed on the left. The VP2 domains and subdomains are represented by a line above the sequence and are colored as in Fig. 3-1D, and the VP1-VP2 interaction sites predicted by Estrozi et al. (EC) is outlined in a box. Amino acid positions are indicated on the right.

Appendix D. Table of Recombinant VP1 and VP2 Mutants. Page 1 of 3

Protein	Strain	Name	Mutated Residue(s)	Notes
VP1	SA11	PDC1	N547A	No loss of function in presence of SA11 VP2 (co-op) compared to WT (Chapter 4)
VP1	SA11	PDC1	M548A	No loss of function in presence of SA11 VP2 (co-op) compared to WT (Chapter 4)
VP1	SA11	PDC1	T549A	No loss of function in presence of SA11 VP2 (co-op) compared to WT (Chapter 4)
VP1	SA11	PDC1	N550A	No loss of function in presence of SA11 VP2 (co-op) compared to WT (Chapter 4)
VP1	SA11		T644A	Low expression in Sf9 cells
VP1	SA11		N640A	No loss of function in presence of SA11 VP2 (co-op) compared to WT
VP1	SA11	PDC2	R614A	Significant loss of function in presence of SA11 VP2 (co-op) compared to WT (Chap. 4)
VP1	SA11	PDC2	N617A	No loss of function in presence of SA11 VP2 (co-op) compared to WT (Chapter 4)
VP1	SA11	PDC2	K618A	No loss of function in presence of SA11 VP2 (co-op) compared to WT (Chapter 4)
VP1	SA11	PDC2	S616A/ S620A	No loss of function in presence of SA11 VP2 (co-op) compared to WT (Chapter 4)
VP1	SA11	PDC2	R614A/S616A/ N617A/K618A/ S620A	Low expression in Sf9 cells (Chapter 4)
VP1	SA11	EC2	P968A	No loss of function in presence of SA11 VP2 (co-op) compared to WT (Chapter 4)
VP1	SA11	EC2	I970A	No loss of function in presence of SA11 VP2 (co-op) compared to WT (Chapter 4)
VP1	SA11	EC2	S978A	No loss of function in presence of SA11 VP2 (co-op) compared to WT (Chapter 4)
VP1	SA11	EC2	D971A	Low expression in Sf9 cells
VP1	SA11	EC2	K969A/D973A/ T974A	No loss of function in presence of SA11 VP2 (co-op) compared to WT (Chapter 4)
VP1	SA11	EC2	S978A/I980A	No loss of function in presence of SA11 VP2 (co-op) compared to WT (Chapter 4)

Appendix D. Table of Recombinant VP1 and VP2 Mutants. Page 2 of 3

Protein	Strain	Name	Mutated Residue(s)	Notes
VP1	SA11	EC2	D971A/S978A/ I980A	Significant loss of function in presence of SA11 VP2 (co-op) compared to WT (Chap. 4)
VP1	SA11	EC2	D971A/Y975A/ S978A/I980A	Low expression in Sf9 cells Significant loss of function in presence of SA11 VP2 (co-op) compared to WT (Chap. 4)
VP1	SA11		N264A	Glycerol stock in -80, BV not made
VP1	SA11		E266A	Glycerol stock in -80, BV not made
VP1	SA11	EC1	E265A	No loss of function in presence of SA11 VP2 (co-op) compared to WT (Chapter 4)
VP1	SA11	EC1	E265A/L267A	Significant loss of function in presence of SA11 VP2 (co-op) compared to WT (Chap. 4)
VP1	SA11		K1025A	Glycerol stock in -80, BV not made
VP1	SA11	EC3	F1022A/ G1024A	No loss of function in presence of SA11 VP2 (co-op) compared to WT (Chapter 4)
VP1	SA11	EC3	F1022A/K1023 A/G1024A	No loss of function in presence of SA11 VP2 (co-op) compared to WT (Chapter 4)
VP1	SA11	WT		Published in McDonald and Patton 2011
VP1	DS1	WT	co-op for Sf9	Published in Steger et al 2019
VP1	WA	WT	co-op for Sf9	Published in Steger et al 2019
VP1	ETD	WT	co-op for Sf9	Published in Steger et al 2019
VP1	PO-13	WT		Published in McDonald and Patton 2011
VP1	PO-13		+ 3’ terminal RGSRGS link	Published in Steger et al 2019
VP1	Bristol	WT		Published in McDonald and Patton 2011
VP1	SA11	Av:Si	PO-13 9-320	Not expressed in Sf9 cells
VP1	SA11	R456	PO-13 residues 47-106, 146-171, 293-320	Not soluble in current purification protocol
VP1	SA11	PRE1	PO-13 474-547	Published in Steger et al 2019
VP1	SA11	PRE 1+2	PO-13 474-547, 886-1016	Published in Steger et al 2019
VP1	SA11	R123	PO-13 474-547, 587-700, 886-1016	Not active in the presence of SA11 VP2 (co-op) or PO-13 VP2
VP1	SA11	PO-13 NTD+ R123	PO-13 9-320, 474-547, 587-700, 886-1016	Not expressed in Sf9 cells

Appendix D. Table of Recombinant VP1 and VP2 Mutants. Page 3 of 3

Protein	Strain	Name	Mutated Residue(s)	Notes
VP1	SA11	R123 456	PO-13 47-106, 146-171, 293-320, 474-547, 587-700, 886-1016	Not expressed in Sf9 cells
VP1	SA11	VL2	PO-13 306-320	Published in Steger et al 2019
VP1	SA11	VL1	PO-13 149-161	Published in Steger et al 2019
VP1	SA11	Motif D	PO-13 642-668	Low expression in Sf9 cells
VP2	SA11	WT		Published in McDonald and Patton 2011
VP2	SA11		co-op for Sf9	Chapter 4
VP2	DS1	WT		Published in Steger et al 2019
VP2	WA	WT		Published in Steger et al 2019
VP2	ETD	WT		Published in McDonald and Patton 2011
VP2	PO-13	WT		Published in McDonald and Patton 2011
VP2	Bristol	WT		Published in McDonald and Patton 2011
VP2	SA11 NTD	Si:Av	PO-13 residues 1-103	Published in Steger et al 2019
VP2	PO-13 NTD	Av:Si	SA11 residues 1-103	Published in Steger et al 2019
VP2	SA11 NTD	Si:Br	Bristol residues 1-103	Published in McDonald and Patton 2011
VP2	Bristol NTD	Br:Si	SA11 residues 1-103	Published in McDonald and Patton 2011
VP2	SA11		C628A	Total loss of function in presence of SA11 co-op and PO-13 VP2s
VP2	SA11		C1138A	Some loss of function in presence of SA11 and PO-13 VP1s
VP2	SA11		C1285A	Significant loss of function in presence of SA11 and PO-13 VP1s
VP2	SA11		C1684A	Significant loss of function in presence of SA11 and PO-13 VP1s
VP2	SA11		C1759A	Some loss of function in presence of SA11 and PO-13 VP1s
VP2	SA11	Cys-less	C628A/C1138A/ C1285A/C1684A/ C1759A	Total loss of function in presence of SA11 and PO-13 VP1s

Appendix E. Genetic Determinants restricting the reassortment of heterologous NSP2 genes into the simian rotavirus SA11 genome. Page 1 of 12

CPL (now known as CLS) is a co-author on this published manuscript. CPL performed replicase assays to test *in vitro* functionality of chimeric RNAs, made NSP2 homology models, and reviewed/edited the manuscript.

SCIENTIFIC REPORTS

OPEN

Genetic determinants restricting the reassortment of heterologous NSP2 genes into the simian rotavirus SA11 genome

Received: 16 March 2017
Accepted: 6 July 2017
Published online: 24 August 2017

Rebecca Mingo[1], Shu Zhang[1], Courtney P. Long[1,2], Leslie E. W. LaConte[1] & Sarah M. McDonald[1,3]

Rotaviruses (RVs) can evolve through the process of reassortment, whereby the 11 double-stranded RNA genome segments are exchanged among strains during co-infection. However, reassortment is limited in cases where the genes or encoded proteins of co-infecting strains are functionally incompatible. In this study, we employed a helper virus-based reverse genetics system to identify NSP2 gene regions that correlate with restricted reassortment into simian RV strain SA11. We show that SA11 reassortants with NSP2 genes from human RV strains Wa or DS-1 were efficiently rescued and exhibit no detectable replication defects. However, we could not rescue an SA11 reassortant with a human RV strain AU-1 NSP2 gene, which differs from that of SA11 by 186 nucleotides (36 amino acids). To map restriction determinants, we engineered viruses to contain chimeric NSP2 genes in which specific regions of AU-1 sequence were substituted with SA11 sequence. We show that a region spanning AU-1 NSP2 gene nucleotides 784–820 is critical for the observed restriction; yet additional determinants reside in other gene regions. *In silico* and *in vitro* analyses were used to predict how the 784–820 region may impact NSP2 gene/protein function, thereby informing an understanding of the reassortment restriction mechanism.

A shared feature of viruses with segmented RNA genomes is their capacity to undergo reassortment during host cell co-infection[1]. When two or more viral strains co-infect the same cell, they can exchange their genome segments. This process results in reassortant progeny with genes derived from multiple parental strains. In some cases, the new combination of genes confers a selective advantage to the virus, for instance by allowing it to evade the host immune response or infect a new host species[1–4]. Because of its substantial impact on viral evolution, reassortment is thought to occur frequently for segmented RNA viruses, such that heterogeneous pools of viruses exist with seemingly random gene combinations. However, the prevalence of reassortants in most viral populations tends to be much lower than what is expected based upon chance combination alone, and some strains are incapable of reassorting genome segments with each other even during experimental co-infection of cultured cells or animals. These observations suggest that reassortment requires genetic and biochemical compatibility. Therefore, nucleotide and/or amino acid differences between parental (i.e., co-infecting) strains may function as determinants of reassortment restriction. Studies seeking to identify and characterize these determinants for a given segmented RNA virus will not only enhance an understanding of the factors tempering or promoting its evolution, but they will also shed light on the basic biology of its replication cycle.

Rotaviruses (RVs) are 11-segmented, double-stranded (ds) RNA viruses that infect a wide array of mammalian and avian hosts[5]. More than 8 genetically-divergent species of RVs have been identified in nature (groups A–H, and tentative group I), though the vast majority of human disease is caused by group A strains[5–8]. While inter-species reassortment (e.g., between group A and B strains) has never been documented, it was generally assumed that reassortment occurs readily among group A strains[9]. However, this notion has been challenged in the past decade, mainly by the results of large-scale comparative genomic studies of wildtype group A human

[1]Virginia Tech Carilion School of Medicine and Research Institute, Roanoke, VA, USA. [2]Translational Biology, Medicine, and Health Graduate Program, Roanoke, VA, USA. [3]Department of Biomedical Sciences and Pathobiology, Virginia-Maryland College of Veterinary Medicine, Blacksburg, VA, USA. Rebecca Mingo and Shu Zhang contributed equally to this work. Correspondence and requests for materials should be addressed to S.M.M. (email: mcdonasm@wfu.edu)

Appendix E. Genetic Determinants restricting the reassortment of heterologous NSP2 genes into the simian rotavirus SA11 genome. Page 2 of 12

and animal strains[10–21]. Together, these studies demonstrate that group A RVs maintain very specific and stable genome segment constellations (i.e., gene sets) despite co-circulating in the same communities and even co-infecting the same host. For instance, group A human strains generally belong to one of two major genogroups (Wa-like or DS-1-like) or to the minor AU-1-like genogroup[22]. Inter-genogroup reassortants are rarely identified in clinical fecal specimens from normal, otherwise healthy children presenting symptoms of gastroenteritis. Reassortants between group A human and animal strains are also rarely identified in clinical specimens, though there are reports of such strains isolated sporadically from humans, often those with confounding medical conditions, such as an immune-compromised status[2]. Low co-infection rates could certainly contribute to the low frequency of reassortants in nature. Genetic and biochemical incompatibilities between the genes (or encoded proteins) of parental strains also plays an important role. In support of this notion, the number of reassortants is consistently lower than what is expected based upon unrestricted combination following experimental co-infection of cultured cells with group A human and animal strains, and there is a non-random association of genome segments in the reassortants that do arise[23–25]. Thus, nucleotide and/or amino acid differences between group A parental strains are putative determinants of reassortment restriction.

Each of the 11 dsRNA genome segments of RV is organized as a central open-reading frame (ORF) flanked by 5′ and 3′ untranslated regions (UTRs)[5]. The mechanism of reassortment for RV is a natural by-product of the manner in which the genome segments are selectively packaged during the viral replication cycle. While this process is poorly understood, the evidence to date is most consistent with a model wherein one of each of the 11 genome segments is packaged into an assembling virion as a single-stranded, positive-sense RNA (+RNA) precursor[26]. More specifically, the 11 distinct RV +RNAs are thought to engage with each other via RNA-RNA interactions to form an assortment complex. The *cis*-acting packaging signals required for these selective RNA-RNA interactions are predicted to reside in the 5′ and 3′ termini of each segment but have not been determined in detail[27,28]. During or subsequent to encapsidation of the assortment complex, the 11 +RNAs are converted into 11 dsRNA genome segments by the VP1/VP2 viral polymerase complex[29]. Reassortment, therefore, occurs when the +RNAs of one parental strain are displaced by cognate +RNAs of another parental strain in the assortment complex. Given this mechanism, it is not surprising that strain-specific nucleotide differences in the genome segments (i.e., +RNAs) of co-infecting parental strains might restrict reassortment at the level of assortment complex formation, encapsidation, or dsRNA synthesis. Such direct restrictions on reassortment within the co-infected cell may explain why strains belonging to different RV groups do not seem to reassort with each other. There is little to no inter-group sequence homology at the 5′ and 3′ termini of genome segments, which may contribute to failure of RNA-RNA interactions during assortment complex formation. However, for the more closely-related group A RVs with conserved 5′ and 3′ termini, reassortment restriction is more likely to occur in a different, indirect manner. Most group A RV reassortants are likely generated during co-infection, but they would not emerge at appreciable levels in the viral population if they contain gene/protein sets that interact sub-optimally and function poorly together.

In the current study, we sought to investigate determinants of reassortment restriction by employing a previously-described, helper virus-based reverse genetic system[30,31]. This system allows for the targeted reassortment of the NSP2 genome segment (hereafter referred to as the NSP2 gene) into the genetic backbone of simian RV strain SA11. Using this approach, we rescued SA11 reassortants containing NSP2 genes from human RVs Wa or DS-1. However, we could not rescue an SA11 reassortant containing an NSP2 gene from the human strain AU-1, which differs from that of SA11 by 183 nucleotides (36 amino acids). A gain-of-function chimeric mutagenesis approach was used to show that at least some of the key restriction determinants map to a region of the NSP2 gene ORF spanning nucleotides 784–820. We show that the 12 nucleotide differences in this region did not alter the predicted folding of the NSP2 + RNA nor did they affect its capacity to serve as an efficient replication template *in vitro*. Yet, the 7 amino acid differences in this region were predicted to cause changes in NSP2 protein structural dynamics. Altogether, the results of this manuscript inform an understanding of RV reassortment restriction determinants and functional domains within the NSP2 + RNA/protein.

Results and Discussion

Rescue of simian RV reassortants with NSP2 genes derived from human RVs Wa and DS-1. To investigate determinants of RV reassortment restriction, we employed a helper virus-based reverse genetics system that was developed by Trask *et al.*[30] (Fig. 1). Specifically, we tested whether simian RV strain SA11 reassortants could be rescued with NSP2 genes derived from human RV genogroup prototype strains Wa, DS-1, or AU-1 (Fig. 2a). Monkey kidney Cos-7 cells were transfected with plasmids expressing the cloned human RV NSP2 genes as +RNAs (Fig. 1). The cells were then infected with simian RV strain SA11-*ts*E for 24 h at 31 °C to allow the cDNA plasmid-derived +RNAs to reassort into the viral genome. The cloned NSP2 +RNAs were engineered to be resistant to the g8D small hairpin (sh) RNA, and they encode a non-temperature-sensitive NSP2 protein. In contrast, the SA11-*ts*E NSP2 + RNA is g8D shRNA-sensitive and encodes a protein that is non-functional at high temperature[32]. Recombinant reassortant viruses generated in Cos-7 cells were selected for by two rounds of growth at 39 °C in g8D shRNA-expressing monkey kidney MA104 cells (Fig. 1). Plaque assays were then performed to isolate individual viral clones, and % rescue efficiency was calculated from the ratio of the number of clones identified as recombinant over the total number of clones investigated by RT-PCR and nucleotide sequencing.

We found that recombinant SA11 reassortants containing the wildtype SA11 NSP2 gene ($rSA11_{SA11}$) and those containing the NSP2 genes from human RVs Wa or DS-1 ($rSA11_{Wa}$ and $rSA11_{DS\text{-}1}$, respectively) were rescued with 85% efficiency (Fig. 2a). In contrast, we were unable to rescue an SA11 reassortant with the human RV AU-1 NSP2 gene ($rSA11_{AU\text{-}1}$), despite screening 100 plaques (Fig. 2a). The reassortant clones that we did rescue ($rSA11_{SA11}$, $rSA11_{Wa}$, and $rSA11_{DS\text{-}1}$) were triple plaque-purified, amplified by two rounds of growth in MA104 cells, and their identities were confirmed by electropherotyping and complete NSP2 gene sequencing (Fig. 2b and

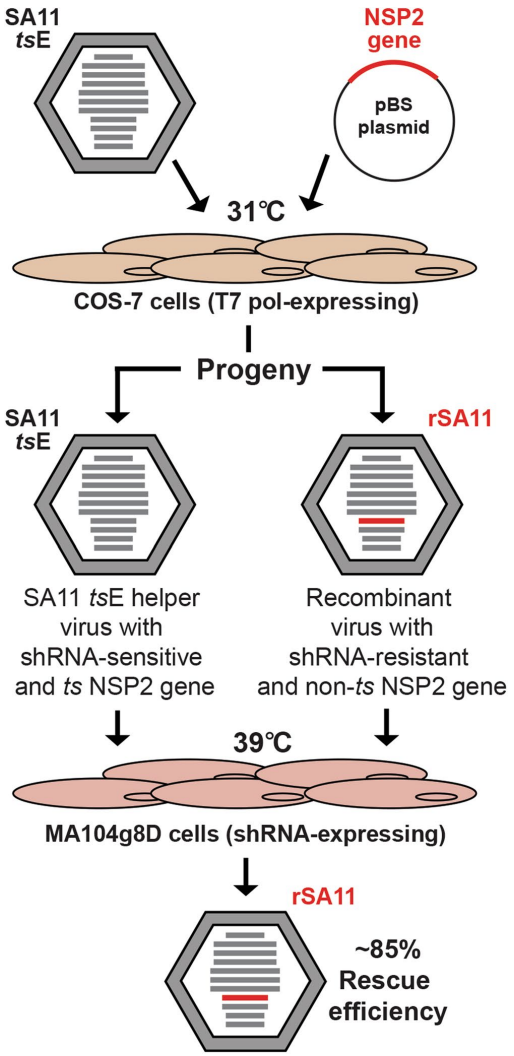


Figure 1. Overview of helper virus-based reverse genetics for the NSP2 gene. Cos-7 cells expressing T7 polymerase were transfected with the pBS plasmid that encodes an NSP2 + RNA. Following infection of the cells with the temperature-sensitive (*ts*) helper RV (SA11-*ts*E) at 31 °C, the plasmid-derived NSP2 + RNA (red) is reassorted into some of the viral progeny, making them recombinant (rSA11). Recombinants viruses are engineered to have an NSP2 gene that is resistant to shRNA degradation and encodes a non-temperature sensitive (*ts*) protein. Following two rounds of growth at 39 °C in shRNA-expressing MA104g8D cells, growth of SA11-*ts*E is suppressed and rSA11 is selected for with an efficiency of ~85% for the wildtype control.

data not shown). The sequences of the NSP2 genes from the rescued viruses were identical to the cDNA clones, suggesting that the recombinant viruses were genetically stable for at least ~5 passages (data not shown). To test whether $rSA11_{Wa}$ or $rSA11_{DS\text{-}1}$ exhibit replication defects, single cycle growth kinetic assays were performed in MA104 cells (Fig. 2c). The results show that both $rSA11_{Wa}$ and $rSA11_{DS\text{-}1}$ replicated as well as the isogenic control virus, $rSA11_{SA11}$, with no statistically-significant differences in peak titers (Fig. 2d). When compared to the SA11 NSP2 gene, that of Wa differs by 198 nucleotides (41 amino acids) and that of DS-1 differs by 225 nucleotides (45 amino acids) (Fig. S1). Based on the efficient rescue and growth of $rSA11_{Wa}$ and $rSA11_{DS\text{-}1}$, we conclude that these differences do not represent reassortment restriction determinants into the SA11 genetic backbone.

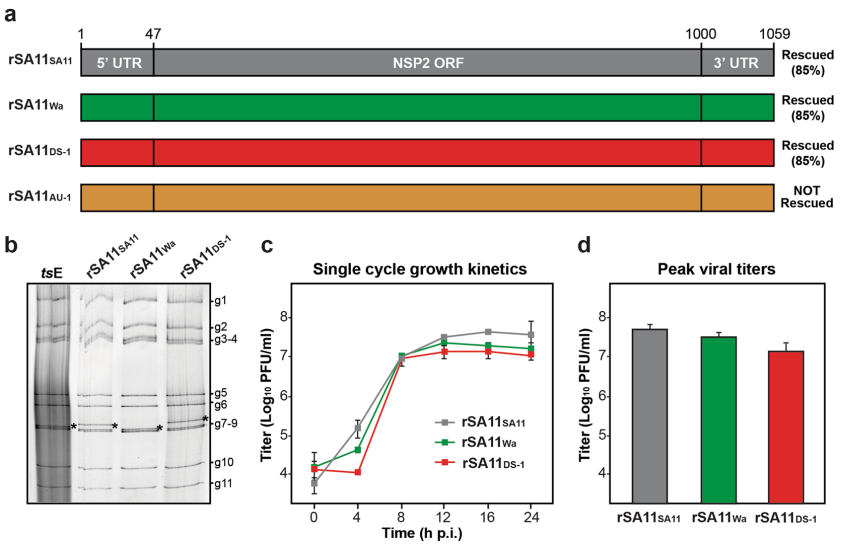


Figure 2. SA11 reassortants with human RV NSP2 genes. (**a**) NSP2 genes of recombinant reassortant viruses $rSA11_{SA11}$, $rSA11_{Wa}$, $rSA11_{DS-1}$, and $rSA11_{AU-1}$ are shown as boxes (not drawn to scale) and colored grey, green, red, and orange, respectively. Nucleotide numbers are listed at the top and the general locations of untranslated regions (UTRs) and open-reading frame (ORF) in the NSP2 gene are shown. The results of the rescue experiments, including % recombinant plaques, are listed to the right of each gene. (**b**) Electropherotyping. Viral dsRNA was extracted from rescued reassortants $rSA11_{SA11}$, $rSA11_{Wa}$, $rSA11_{DS-1}$ and visualized following SDS-PAGE and SYBER Gold staining. The positions of the 11 viral genome segments are labeled to the right of the gel. The asterisk (*) shows the location of the reassorted NSP2 gene in each virus. (**c**) Single cycle growth kinetic assay. MA104 cells were infected at an MOI of 5 PFU per cell with $rSA11_{SA11}$, $rSA11_{Wa}$, or $rSA11_{DS-1}$ and viral titers were determined at various hours (**h**) p.i. by plaque assay. The results of a single representative experiment are shown. Error bars represent standard deviation from the mean calculated from 3–4 plaque assay wells. (**d**) Peak viral titers. The graph shows average peak viral titers for $rSA11_{SA11}$, $rSA11_{Wa}$, or $rSA11_{DS-1}$. Error bars represent standard deviation from the mean from two experiments (6–8 plaque assay wells). The differences in peak titers of $rSA11_{SA11}$, $rSA11_{Wa}$, or $rSA11_{DS-1}$ were not statistically significant.

However, the lack of $rSA11_{AU-1}$ rescue suggests that (i) the AU-1 NSP2 + RNA was not efficiently packaged/replicated by the SA11-*ts*E helper virus and/or that (ii) the AU-1 NSP2 + RNA/protein is not functionally compatible with the other SA11 +RNAs/proteins during the subsequent *de novo* viral replication that is required for plaque formation.

Mapping reassortment restriction determinants in the AU-1 NSP2 gene. The AU-1 NSP2 gene differs from that of SA11 by 183 nucleotides (36 amino acids) (Fig. S1). To narrow down which of these differences correlate with AU-1 NSP2 gene reassortment restriction into SA11, we employed a gain-of-function chimeric mutagenesis approach. Specifically, we cloned AU-1 NSP2 genes that contained the SA11 5′ and 3′ UTRs as well as SA11 sequence fragments in various ORF regions (Fig. 3a). We tested whether we could rescue SA11 reassortants using these chimeric NSP2 genes. We found that we were unable to rescue reassortant $rSA11_{CH1}$, despite the fact that this NSP2 gene has authentic SA11 UTRs. This result suggests that at least some of the restriction determinants map within the gene ORF. Likewise, we could not rescue $rSA11_{CH2}$, which contains SA11 UTRs as well as SA11 sequence in an ORF region spanning nucleotides 47–481. Thus, the downstream region of 482–1000 must contain important restriction determinants. In support of this notion, we were able to rescue $rSA11_{CH3}$ and $rSA11_{CH4}$ with efficiencies of 85% and 15%, respectively. The reassortant $rSA11_{CH3}$ virus contains an AU-1 NSP2 ORF in which nucleotides 482–1000 were replaced with SA11 sequence. The reassortant $rSA11_{CH4}$ virus contains an AU-1 NSP2 ORF in which only a smaller sub-region (SR) corresponding to nucleotides 784–820 was replaced with SA11 sequence. That rationale for changing only the SR was based upon the observation that 36% of the coding changes clustered in this 37-nucleotide stretch of the NSP2 ORF.

The reassortant clones ($rSA11_{CH3}$ and $rSA11_{CH4}$) were triple plaque-purified, amplified by two rounds of growth in MA104 cells, and their identities were confirmed by electropherotyping and complete NSP2 gene sequencing (Fig. 3b and data not shown). The sequences of the NSP2 genes from the rescued viruses were identical to the cDNA clones, suggesting that the recombinant viruses were genetically stable at least for ~5 passages (data not shown). To test whether $rSA11_{CH3}$ or $rSA11_{CH4}$ exhibit replication defects, single cycle growth kinetic assays were performed in MA104 cells (Fig. 3c). The results show that both $rSA11_{CH3}$ and $rSA11_{CH4}$ had lower

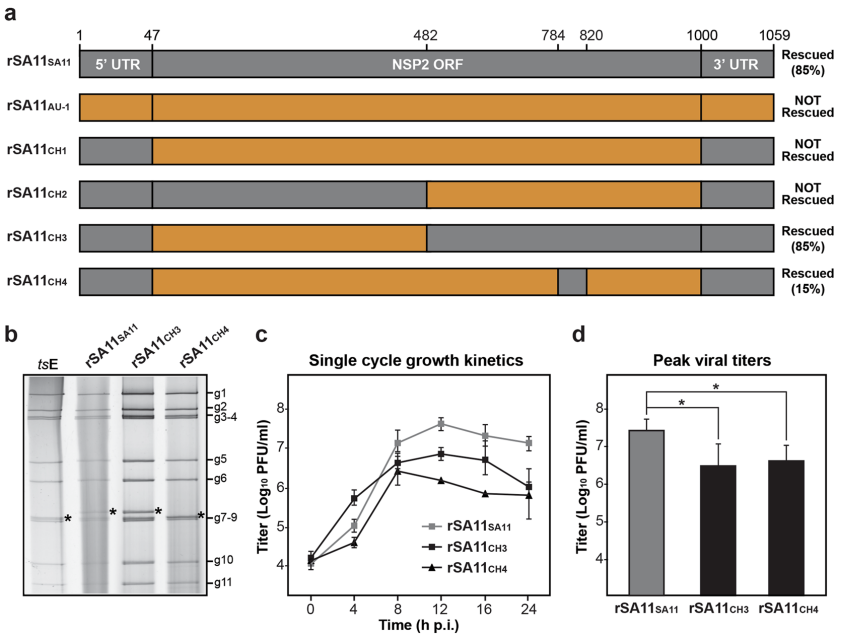


Figure 3. SA11 reassortants with chimeric NSP2 genes. (**a**) NSP2 genes of recombinant reassortant viruses are shown as boxes (not drawn to scale). For all genes, SA11 sequence is shown in grey and AU-1 sequence is shown in orange. Nucleotide numbers are listed at the top and the general locations of untranslated regions (UTRs) and open-reading frame (ORF) in the NSP2 gene are shown. The result of the rescue experiments, including % recombinant plaques, are listed to the right of each gene. (**b**) Electropherotyping. Viral dsRNA from was extracted from rescued reassortants $rSA11_{SA11}$, $rSA11_{CH3}$, $rSA11_{CH4}$ visualized following SDS-PAGE and SYBER Gold staining. The positions of the 11 viral genome segments are labeled to the right of the gel. The asterisk (*) shows the location of the reassorted NSP2 gene in each virus. (**c**) Single cycle growth kinetic assay. MA104 cells were infected at an MOI of 5 PFU per cell with $rSA11_{SA11}$, $rSA11_{CH3}$, or $rSA11_{CH4}$ and viral titers were determined at various hours (**h**) p.i. by plaque assay. The results of a single representative experiment are shown. Error bars represent standard deviation from the mean calculated from 3–4 plaque assay wells. (**d**) Peak viral titers. The graph shows average peak titer for $rSA11_{SA11}$, $rSA11_{CH3}$, or $rSA11_{CH4}$. Error bars represent standard deviation from two experiments (6–8 plaque assay wells). Asterisks (*) indicate statistically significant differences ($p < 0.05$).

titers than the control $rSA11_{SA11}$ virus at several times p.i. and both had statistically-significant lower peak viral titers (approximately 1 $\log_{10}$ step) compared to $rSA11_{SA11}$ (Fig. 3d). This result suggests that both $rSA11_{CH3}$ and $rSA11_{CH4}$ have modest replication defects. However, the observation that that $rSA11_{CH4}$ was rescued at all, despite its replication defect, indicates that the SR (nucleotides 784–820) contains important determinants of AU-1 NSP2 gene reassortment restriction. Still, the higher rescue efficiency of $rSA11_{CH3}$ compared to $rSA11_{CH4}$ (i.e., 85% v. 15%) indicate that additional residues outside the SR but within the larger 482–1000 region are important for robust NSP2 gene reassortment. When comparing the NSP2 gene sequences of SA11 vs. AU-1, we found a total of 95 nucleotide (19 amino acid) differences in the 482–1000 region, and only 12 nucleotide (7 amino acid) differences reside within the SR (Fig. 4). At this time, we cannot exclude the possibility that the UTRs of the AU-1 NSP2 gene also contain restriction determinants when appended to the SA11 ORF. This is especially true for the 3′ UTR of the AU-1 NSP2 gene, which shows 7 nucleotide differences when compared to that of SA11; the 5′UTR only has a single nucleotide change and is unlikely to have a dramatic impact on reassortment restriction (Fig. S1).

Predicting the contributions of the 12 SR nucleotide changes on the NSP2 + RNA. We next sought to gain insight into how the 12 SA11-specific nucleotides in the SR might have increased the efficiency of the AU-1 NSP2 gene reassortment into the SA11 genetic backbone in the context of $rSA11_{CH4}$ allowing rescue using the reverse genetics system. We reasoned that the NSP2 gene + RNA of $rSA11_{CH4}$ (CH4 + RNA) might show differences in its folding or function relative to that of $rSA11_{CH1}$ (CH1 + RNA), which was not rescued in our experiments and lacks the 12 SA11-specific SR nucleotides, but maintains the SA11 UTRs (Fig. 3a). To assess this possibility, we predicted the minimum free energy secondary structures and base pair probabilities of the +RNAs *in silico* (Fig. 5a). The results show that CH1 and CH4 +RNAs are predicted to have similar overall folds with only

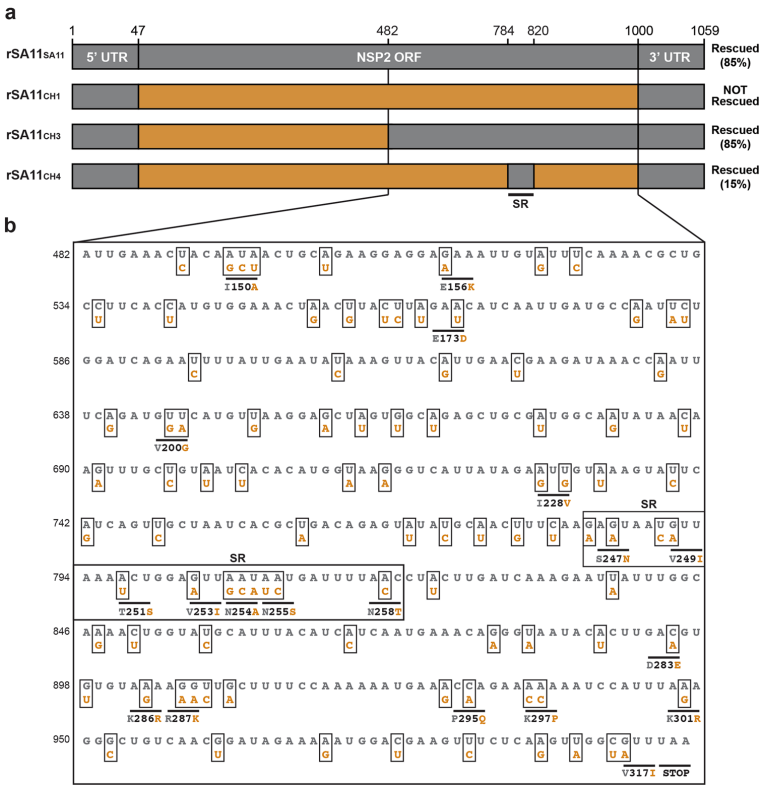


Figure 4. Nucleotide and amino acid differences between SA11 NSP2 and AU-1 NSP2 in the 482–1000 region. (**a**) NSP2 genes of recombinant reassortant viruses with wildtype SA11 or chimeric NSP2 genes are shown as boxes (not drawn to scale). For all genes, SA11 sequence is shown in grey and AU-1 sequence is shown in orange. Nucleotide numbers are listed at the top and the general locations of untranslated regions (UTRs) and open-reading frame (ORF) in the NSP2 gene are shown. The results of the rescue experiments, including % recombinant plaques, are listed to the right of each gene. (**b**) The sequence of nucleotides 482–1000 of SA11 NSP2 are shown in grey. The nucleotides that differ for AU-1 NSP2 are shown below the SA11 sequence and are colored orange. Amino acid changes are listed beneath the corresponding codon with the SA11 residue shown in grey and the AU-1 residue shown in orange. The SR (nucleotides 784–820) is outlined with a box and labeled.

minor energetic changes in the SR. To test whether CH1 and CH4 +RNAs have differences in their capacity to serve as templates for replication, we employed an *in vitro* dsRNA synthesis assay (Fig. 5b). We found that CH1 and CH4 NSP2 +RNAs were efficiently converted into dsRNA products by SA11 VP1/VP2 at levels comparable to the wildtype SA11 NSP2 + RNA control. Thus, the NSP2 + RNA of the non-rescued $rSA11_{CH1}$ virus is indistinguishable in its predicted folding and *in vitro* replicative capacity from that of the rescued $rSA11_{CH4}$ virus. This result suggests that the difference in rescue efficiency/viral growth between $rSA11_{CH1}$ and $rSA11_{CH4}$ cannot be easily explained at the + RNA level. Still, it remains possible that the CH1 + RNA has defects in RNA-RNA interactions required for assortment/packaging into SA11, which we are unable to experimentally test using this helper virus-based reverse genetics system.

Predicting the contributions of 7 SR amino acid changes on NSP2 protein structural dynamics. NSP2 is a 35-kDa viral protein that forms a stable, donut-shaped octamer[32,33]. NSP2 interacts with viral RNA and four viral proteins (VP1, VP2, NSP5 and NSP6), and it is essential for RV particle assembly and dsRNA synthesis[34–39]. One possible reason for the lack of rescuable AU-1 and CH1 NSP2 gene reassortants in our reverse genetics experiments could be that the AU-1 NSP2 protein does not functionally interact or interacts only sub-optimally with its viral binding partners (or with as of yet unknown cellular binding partners). In this case, even if the reassortment event occurred in the transfected-infected Cos-7 cells, the $rSA11_{AU-1}$ and $rSA11_{CH1}$ viruses would not replicate to detectable levels in MA104g8D cells. By changing the 7 AU-1-specific amino acid residues of the SR to match those of SA11 in the context of $rSA11_{CH4}$, we may have restored the capacity of the

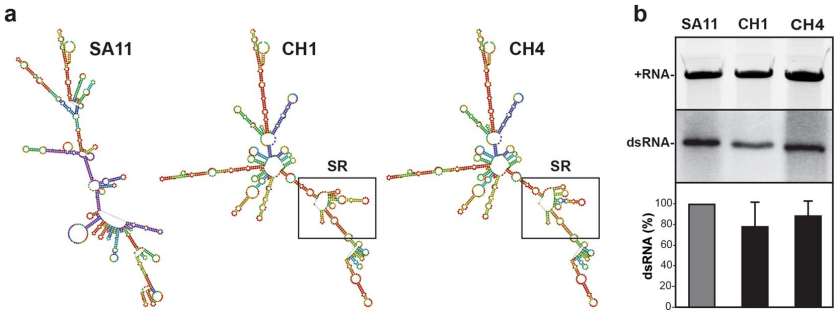


Figure 5. Predicted folding and *in vitro* replication of NSP2 +RNAs. (**a**) The predicted secondary structures of SA11, CH1, and CH4 +RNAs are shown and colored according to base-pair probabilities (warmer color = more favorable, cooler color = less favorable). The nucleotides comprising the SR of CH1 and CH4 + RNA is outlined with a box. (**b**) *In vitro* replicase assays. The SA11, CH1, and CH4 + RNA templates were electrophoresed in a 7M-urea-acrylamide gel and visualized following ethidium bromide staining (top gel panel). Gel image was cropped to show only the + RNA bands. These +RNAs were then assayed for their capacity to serve as templates for dsRNA synthesis by SA11 VP1/VP2 *in vitro*. The nascent minus-strands of the dsRNAs were radiolabeled using [^{32}P]αUTP and detected following SDS-PAGE and autoradiography (bottom gel panel, representative experiment). Gel image was cropped to show only the dsRNA bands. The dsRNA levels were quantified for 4 independent experiments and shown as a % relative to SA11 dsRNA. The differences in CH1 and CH4 dsRNA levels relative to SA11 are not statistically significant.

NSP2 protein to better interact with its binding partners, thereby improving replication enough to rescue the virus.

To gain insight into possible differences in the NSP2 protein as a result of the 7 SR amino acid changes, we performed unrestrained molecular dynamics simulations. Specifically, the monomeric structures of SA11 NSP2, AU-1 NSP2, and CH4 NSP2 were each simulated for 20 nanoseconds at 300 K. Root mean square fluctuations (RMSFs) of the α-carbons were used to calculate B-factors, thereby providing an estimate of the degree of flexibility at each position of the protein. Several regions of flexibility were predicted in all three NSP2 proteins (Fig. 6a). In particular, the SR (positions 247–258) is a flexible loop element, and it is followed by a flexible C-terminal region (CTR; positions Change "258" to "275"–311). The extreme C-terminus (positions 312–317) showed dramatic, but consistent, movement for all three protein structures. Movement of residues in and around the SR (i.e., positions 235–243 and 251–256) have been described to occur following RNA binding of NSP2[34]. Moreover, flexibility of the NSP2 C-terminus has been previously reported, and it is suggested to be important for domain-swapping interactions that link NSP2 octamers together[37].

We reasoned that the structural dynamics of CH4 NSP2 would largely match those of AU-1 NSP2, except at sites that were functionally restored due to the 7 SR amino acid changes. We thought that those restored sites of CH4 NSP2 would show flexibility similar to that observed for SA11 NSP2. In particular, we predicted that the relative flexibility of the SR would be the same for CH4 NSP2 and SA11 NSP2 because they are the same sequence. Surprisingly, we found that most SR positions exhibited lower B-factors for CH4 NSP2 compared to both SA11 and AU-1 NSP2 proteins (Fig. 6b). The flexibility at only SR positions 255–258 were similar between CH4 NSP2 and SA11 NSP2, but differed for that of AU-1 NSP2 (Fig. 6b). We also identified 12 positions in the CTR (278–279, 293–296, 301, and 305–309) for which CH4 NSP2 flexibility matched that of SA11 NSP2, but not AU-1 NSP2 (Fig. 6c). While general trends in the data were observed, only positions 305 and 306 showed statistically significant differences in B-factors for both SA11 v. AU-1 NSP2 as well as AU-1 v. CH4 NSP2. Still, the decreased B-factors observed at all 16 positions in AU-1 NSP2, which did not allow for rescue of an SA11 reassortant, are indicative of a more ordered structure that is potentially restricted in interacting with binding partners or undergoing conformational change. Thus, the restoration of motion in CH4 NSP2 in the SR and CTR may be essential for optimal NSP2 interactions with binding partners, thereby contributing to the rescue of the rSA11$_{CH4}$ virus.

Using the homology model of CH4 NSP2 (Fig. 7a), we determined the three-dimensional location of the 16 positions that exhibited restored flexibility. These 16 positions comprise a large surface-exposed patch on the side of the octamer (Fig. 7b). The location of this patch is amenable to protein/RNA interactions, though it does not significantly overlap with any of the known or predicted binding sites on NSP2[34–39]. The molecular dynamics simulations and modeling results suggest that that amino acid changes in the NSP2 SR can alter the structural dynamics of the NSP2 CTR. Thus, this data raises questions about whether altering amino acids as a result of the SR swap alleviated proximal structure-based determinants of gene reassortment restriction for AU-1 NSP2 into SA11.

Summary and conclusions. It has long been observed that reassortment is restricted among some group A RV strains; yet the genetic determinants that mediate such restriction are completely unknown. In the current study, we employed a helper virus-based reverse genetics system and attempted to engineer simian RV strain

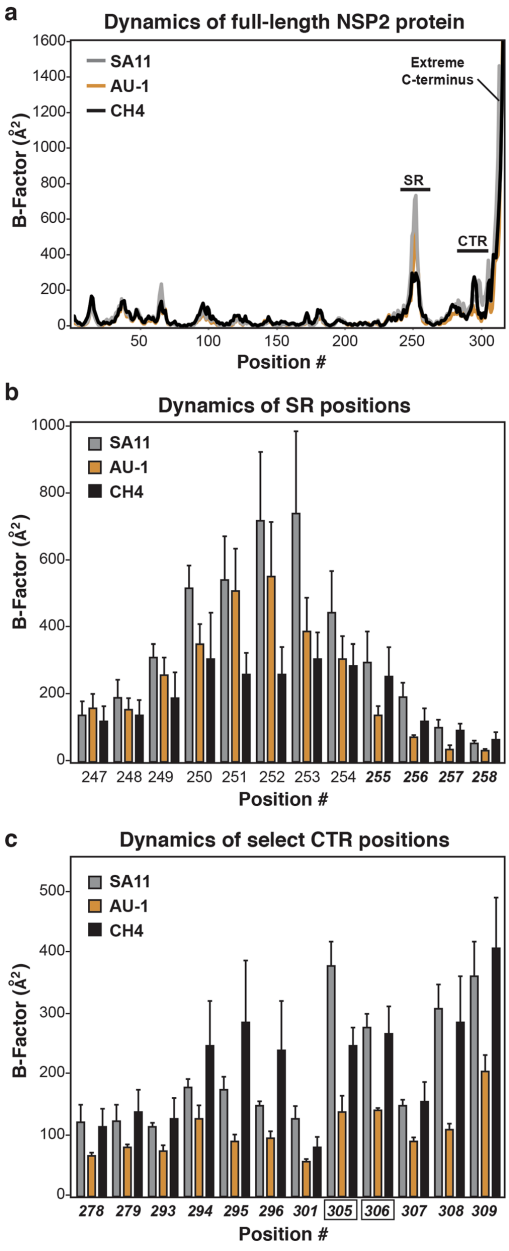


Figure 6. Molecular dynamics simulations of NSP2 proteins. (**a**) Average B-factors are shown for each position of SA11 NSP2 (grey), AU-1 NSP2 (orange), and CH4 NSP2 (black). The regions corresponding to the SR, CTR, and extreme C-terminus of the NSP2 protein are labeled. Dynamics of the (**b**) SR positions and (**c**) select CTR positions. The graphs indicate relative flexibility at individual positions for SA11 NSP2 (grey), AU-1 NSP2 (orange), and CH4 NSP2 (black). Error bars represent standard deviation from the mean following 3 independent simulations. The positions numbers in bold-italic type are those in which the predicted B-factors of CH4 NSP2 are similar to those of SA11 NSP2. Positions 305 and 306 (black box outline) showed statistically-significant differences ($p < 0.05$) when comparing B-factors for SA11 v. AU-1 NSP2 and AU-1 v. CH4 NSP2.

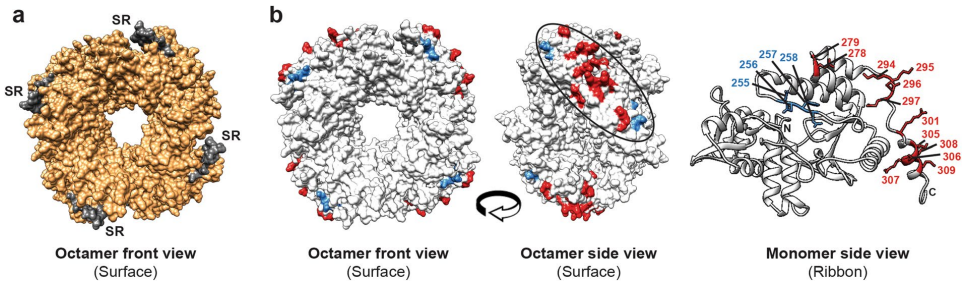


Figure 7. Location of dynamically-restored residues on the CH4 NSP2 protein structure. (**a**) Homology model of CH4 NSP2 is shown as an octamer (surface-filled and viewed from the front). Regions of the protein corresponding to AU-1 NSP2 sequence are shown in orange and those of SA11 sequence (i.e., the SR) are shown in grey. (**b**) Left: The CH4 NSP2 octamer is shown in the same position as panel A but colored white. The 4 restored positions that map to the SR are shown in blue and the 12 restored positions that map to the CTR are shown in red. Middle: These 16 positions form a large patch on the side of the octamer structure. Right: The restored positions are shown in stick representation and labeled on a single CH4 NSP2 monomer ribbon structure.

SA11 reassortants containing NSP2 genes from prototypic human RV strains Wa, DS-1, and AU-1 ($rSA11_{Wa}$, $rSA11_{DS-1}$, and $rSA11_{AU-1}$, respectively). We showed that $rSA11_{Wa}$ and $rSA11_{DS-1}$ were efficiently rescued and exhibited no detectable growth defects. In contrast, $rSA11_{AU-1}$ was not rescuable in this system. This result was surprising and led us to test whether we could overcome the observed reassortment restriction by replacing regions of the AU-1 NSP2 gene (i.e., that contain restriction determinants) with SA11 sequence (i.e., that lack restriction determinants). Using this approach, we found that restriction determinants map to the SR (nucleotides 784–820) within the AU-1 NSP2 gene ORF. However, it is important to note that the SR is not the *only* region containing restriction determinants; there are certainly other regions that contributed to the lack of rescuable $rSA11_{AU-1}$ virus. In support of this notion, we found that $rSA11_{CH3}$ had modest a replication defect, indicating that determinants of robust reassortant viral growth reside within ORF nucleotides 47–481. While not tested here, the 3′UTR of the AU-1 NSP2 gene may also contain restriction determinants. Nevertheless, the molecular details of exactly *how* any of these regions influence reassortment restriction is not understood. For the SR, we hypothesize that the restriction determinants do not prevent packaging of the AU-1 NSP2 + RNA into SA11 during the assortment process in Cos-7 cells. Instead, we think that it is more likely that the AU-1-specific SR nucleotides/amino acids reduced replication of the reassortant virus, thereby preventing its detection following growth in MA104g8D cells. Our *in vitro* and *in silico* analyses of the CH1 vs. CH4 +RNAs, which differ by 12 nucleotides, did not reveal any obvious differences that could account for the restriction being mediated at the RNA-level. In contrast, our *in silico* analyses of the NSP2 protein showed that the SR encodes a highly flexible and surface-exposed loop element. The 7 amino acid changes that resulted from the SR sequence swap are predicted to alter the local flexibility of the AU-1 NPS2 SR and CTR in a way that more closely resembles the that of SA11 NSP2. It is possible, albeit not experimentally validated, that this change in flexibility allowed the chimeric NSP2 protein, but not the AU-1 NSP2 protein, to engage viral or cellular binding partners, improving its replicative fitness. The SR is a highly variable site on the NSP2 protein (<50% sequence identity among group A RV strains). When comparing the SR sequences of strains SA11, Wa, DS-1, and AU-1, we find that AU-1 has unique residues at positions 249, 253–255, and 258, which may be the primary determinants for restriction in this case (Fig. S1). Ongoing and future work in our laboratory is employing biochemical approaches to test the hypothesis that restriction is mediated by sub-optimal NSP2-protein or NSP2-RNA interactions due to the identified amino acid changes.

During the writing of this manuscript, a plasmid only-based reverse genetics system was reported by Kanai *et al*., which allows for the targeted manipulation of any SA11 gene and rescue in the absence of a helper virus[40]. Without a doubt, this new system will provide an improved experimental platform for detailed investigations of RV gene reassortment restriction mechanisms, including the study of genetic epistasis among multiple segments. However, one advantage of the helper virus-based system that we used here is that it more closely mimics co-infection conditions. Much like in nature, only reassortants with at least a moderate degree of fitness will be rescued using the helper virus-based approach, because it requires that recombinant +RNAs successfully compete with those of the helper virus. It is possible that replication-defective reassortants will be rescued using the plasmid-only reverse genetic system because there is no competition of wildtype +RNAs during the assortment step; these same reassortants would not emerge from co-infected cells. On the other hand, the fully plasmid-based system will allow for the differentiation of direct restrictions (i.e., failure of + RNA packaging) vs. indirect restrictions (i.e., failure of +RNAs/proteins to interact during *de novo* replication). In essence, non-rescuable reassortants from the fully-plasmid system must represent a direct restriction that prevents the generation of hybrid progeny. Regardless of the genetic system used, studies seeking to uncover the molecular details that prevent

reassortants from emerging in the RV population are warranted because they will inform rational infection control measures and elucidate new aspects of virus biology.

Materials and Methods

Cells and viruses. Monkey kidney cell lines (MA104 and Cos-7) were obtained from American Type Cell Culture and maintained as described by Arnold *et al.*, 2009 in medium (Life Technologies) that was supplemented to contain 5–10% heat-inactivated fetal bovine serum, 100 U/ml penicillin, 100 µg/ml streptomycin, and 0.5 µg/ml Amphotericin B[41]. MA104g8D cells stably expressing a short hairpin RNA (shRNA) against the NSP2 gene were obtained from Dr. John Patton (University of Maryland, College Park) and cultured in the same manner as MA104 cells[30]. RV strains Wa, DS-1, and AU-1 were obtained from Dr. John Patton (University of Maryland, College Park). RV strain SA11-*ts*E (clone 1400) was obtained from Dr. John Patton with permission from Dr. Frank Ramig (Baylor College of Medicine)[42]. RV infections were performed as described previously[41]. Replication-defective vaccinia virus expressing T7 polymerase (strain rDI-T7pol) was obtained from Dr. John Patton (University of Maryland, College Park) with permission from Dr. Koki Taniguchi (Fujita Health University School of Medicine, Japan) and was propagated as described previously[31,32].

Plasmid construction. A plasmid expressing the full-length, shRNA-resistant NSP2 gene from simian RV strain SA11 (pBS-SA11g8^R) was obtained from Dr. John Patton (University of Maryland, College Park)[31]. To construct similar plasmids expressing the full-length NSP2 genes from human RV strains Wa (pBS-Wag8^R), DS-1 (pBS-DS-1g8^R), or AU-1 (pBS-AU-1g8^R) RT-PCR and molecular cloning was used. Specifically, viral RNA was extracted from MA104 cells infected with Wa, DS-1, or AU-1 and used as template in RT-PCR reactions to produce cDNA. Primer-generated restrictions sites 5′ *NcoI* and 3′ *AfeI* allowed the cDNA products of the Wa, DS-1, or AU-1 NSP2 genes to be individually ligated into the pBS-SA11g8^R plasmid in place of the SA11 NSP2 gene. Site-directed mutagenesis was then used to engineer shRNA resistance in the AU-1 NSP2 gene. The cloned Wa and DS-1 NSP2 gene sequences are identical to what is reported in GenBank (accession numbers RO1NS35F and EF672580, respectively). The AU-1 NSP2 gene sequence is identical to what is reported in GenBank (accession number DQ490534) with the exception of a single G-to-A nucleotide change at position 40 in the 5′ UTR and the 6 nucleotide changes in the shRNA target site (nucleotides 655–673). The NSP2 proteins encoded by all three cloned genes are identical to GenBank sequences. To engineer the chimeric NSP2 genes, PCR was used to amplify specified regions (Fig. 3a) from the pBS-SA11g8^R and pBS-AU-1g8^R plasmids. The Golden Gate Assembly kit (New England Biolabs) was then used to seamlessly ligate the amplified cDNAs together, creating plasmids pBS-CH1g8^R, pBS-CH2g8^R, pBS-CH3g8^R, and pBS-CH4g8^R. Final plasmids were sequenced to ensure integrity of the gene and absence of PCR-induced mutations. All primers used for cloning are available by request.

Generation and recovery of recombinant SA11 reassortants. Recombinant SA11 reassortants were generated as described previously[31,32]. Briefly, Cos-7 cells in 6-cm tissue culture dishes (~1.25×10^6 cells/well) were inoculated with rDI-T7pol at a multiplicity of infection (MOI) of 3 plaque-forming units (PFU) per cell and then transfected with 2.5 µg of plasmid DNA (pBS-SA11g8^R, pBS-Wag8^R, pBS-DS-1g8^R, pBS-AU-1g8^R, pBS-CH1g8^R, pBS-CH2g8^R, pBS-CH3g8^R, or pBS-CH4g8^R) using Trans-IT LT1 (Mirus). The cells were incubated at 37 °C for 24 hours to allow for + RNA expression from the plasmids, and then they were infected with the helper virus SA11-*ts*E at an MOI of 5 PFU per cell at 31 °C for 24 hours. Cells were harvested by 3 rounds of freeze-thaw, and then the clarified supernatant (P0 stock) was passaged in MA104g8D cells at 39 °C for two rounds (P1 and P2 stocks) in the presence of 10 µM AraC (Fisher) to inhibit rDI-T7pol replication[31]. Recombinant reassortants were detected in the P2 stocks by RT-PCR and sequencing of the NSP2 gene. For rSA11-NSP2$_{SA11}$, rSA11-NSP2$_{Wa}$, rSA11-NSP2$_{DS-1}$ and rSA11-NSP2$_{CH3}$, 17/20 plaques each were identified as recombinant (85% rescue efficiency). For rSA11-NSP2$_{CH4}$, 3/20 plaques were identified as recombinant (15% rescue efficiency). For rSA11-NSP2$_{AU-1}$, rSA11-NSP2$_{CH1}$, and rSA11-NSP2$_{CH2}$ no recombinant viruses were detected in the P2 stocks by RT-PCR/sequencing, nor from screening 20–100 individual plaques. Rescued recombinants were subjected to three rounds of plaque-purification in MA104 cells to isolate viral clones[41]. Final high-titer SA11 reassortant stocks were grown for two passages in MA104 cells, and they were verified by RT-PCR, sequencing of the NSP2 gene, and electrophoretic analysis of viral dsRNA as described previously[31].

Single cycle growth kinetic assays. MA104 cells were infected with the recombinant wildtype control virus (rSA11-NSP2$_{SA11}$) or a rescued reassortant (rSA11-NSP2$_{Wa}$, rSA11-NSP2$_{DS-1}$, rSA11-NSP2$_{CH3}$, or rSA11-NSP2$_{CH4}$) at an MOI of 5 PFU per cell. At the indicated times p.i., cells were harvested by 3 rounds of freeze-thaw, and infectious virus in the clarified lysate was quantitated by plaque assay[41]. Titers at 4, 8, 12, 16 and 24 hours p.i. were normalized based upon the titers at 0 hours p.i. for each virus. Two-tailed tests were performed using Smith's Statistical Software, and p values < 0.05 were considered to be statistically significant.

Modeling of NSP2 + RNA folding and *in vitro* replicase assays. Minimum free energy secondary structure and base pair probabilities were calculated for SA11, CH1, and CH4 NSP2 +RNAs using the RNAfold webserver (http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi)[43]. The + RNA templates used in the *in vitro* replicase assays were prepared using the reverse genetic plasmids (pBS-SA11g8^R, pBS-CH1g8^R, and pBS-CH4g8^R) and the protocol described by McDonald *et al.*[44]. Template quantity was determined by using a UV spectrophotometer (optical density at 260 nm [OD_{260}]) and quality was assessed by electrophoresis in 7 M urea-5% polyacrylamide gels stained with ethidium bromide. SA11 VP1 and VP2 purifications and *in vitro* replicase assays were performed as described by McKell *et al.*[45]. Two-tailed tests were performed using Smith's Statistical Software, and p values < 0.05 were considered to be statistically significant.

Appendix E. Genetic Determinants restricting the reassortment of heterologous NSP2 genes into the simian rotavirus SA11 genome. Page 11 of 12

Molecular dynamics simulations of NSP2 proteins. Molecular dynamics simulations using a standard approach[46] were performed using GROMACS v5.1.3[47] on the atomic structure of the SA11 NSP2 (PDB#1L9V) monomer or on homology models of AU-1 NSP2 or CH4 NSP2 monomers, which were created using UCSF Chimera[48]. The PDB files of the modeled structures are available upon request. Prior to performing the simulations, the complete structures were explicitly solvated with a three-point water model (TIP3P) in rhombic dodecahedron water box (solute-box distance of 1.0 nm) under periodic boundary conditions, with charges neutralized by chloride ions. The AMBER99SB-ILDN force field was used for all simulations[49]. Starting structures were energy minimized until convergence at Fmax < 1000 kJ/mol/nm. A 100-picosecond position-restrained NVT equilibration simulation was run for water relaxation at 300 K using a modified Berendsen (velocity rescaling) thermostat, followed by a 100-picosecond NPT equilibration simulation using the Parrinello-Rahman barostat for pressure coupling. After equilibration, an unrestrained 20-nanosecond NPT molecular dynamics simulation was run at 300 K. Three trajectories initiated with different random seeds were run for each protein structure. The RMSF of α-carbons from each of the three trajectories was calculated using the gmx rmsf command in GROMACS. B-factors (i.e., Debye-Waller factors) for each residue were calculated from the RMSF values using an established equation [B-factor $= (8\Pi^2/3) \times (\mathrm{RMSF})^2$][50,51]. One-way ANOVA analyses were performed using StatPlus, and p values < 0.05 were considered to be statistically significant.

Data availability. The datasets generated during and/or analyzed during the current study are available from the corresponding author upon request.

References

1. McDonald, S. M., Nelson, M. I., Turner, P. E. & Patton, J. T. Reassortment in segmented RNA viruses: mechanisms and outcomes. *Nat. Rev. Microbiol.* **14**, 448–460 (2016).
2. Ghosh, S. & Kobayashi, N. Exotic rotaviruses in animals and rotaviruses in exotic animals. *Virusdisease* **25**, 158–172 (2014).
3. Steel, J. & Lowen, A. C. Influenza A virus reassortment. *Curr. Top. Microbiol. Immunol.* **385**, 377–401 (2014).
4. Li, C. & Chen, H. Enhancement of influenza virus transmission by gene reassortment. *Curr. Top. Microbiol. Immunol.* **385**, 185–204 (2014).
5. Estes, M. K. & Kapikian, A. Z. Rotaviruses and Their Replication, In: Knipe, D. M. & Howley, P. M. (Eds). *Fields Virology*. 5th Edition. Lippincott Williams and Wilkens; Philadelphia, p. 1917–1974 (2007)
6. Matthijnssens, J. *et al.* VP6-sequence-based cutoff values as a criterion for rotavirus species demarcation. *Arch. Virol.* **157**, 1177–1182 (2012).
7. Mihalov-Kovács, E. *et al.* Candidate new rotavirus species in sheltered dogs, Hungary. *Emerg. Infect. Dis.* **21**, 660–663 (2015).
8. Bányai, K. *et al.* Candidate new rotavirus species in Schreiber's bats, Serbia. *Infect. Genet. Evol.* **48**, 19–26 (2017).
9. Gouvea, V. & Brantly, M. Is rotavirus a population of reassortants? *Trends. Microbiol.* **3**, 159–162 (1995).
10. Dennis, A. F. *et al.* Molecular epidemiology of contemporary G2P[4] human rotaviruses cocirculating in a single U.S. community: footprints of a globally transitioning genotype. *J. Virol.* **88**, 3789–3801 (2014).
11. Donato, C. M. *et al.* Characterization of G2P[4] rotavirus strains causing outbreaks of gastroenteritis in the Northern Territory, Australia, in 1999, 2004 and 2009. *Infect. Genet. Evol.* **28**, 434–445 (2014).
12. Donato, C. M. *et al.* Characterization of a G1P[8] rotavirus causing an outbreak of gastroenteritis in the Northern Territory, Australia, in the vaccine era. *Emerg. Microbes. Infect.* **3**, e47 (2014).
13. Heiman, E. M. *et al.* Group A human rotavirus genomics: evidence that gene constellations are influenced by viral protein interactions. *J. Virol.* **82**, 11106–11116 (2008).
14. Matthijnssens, J. & Van Ranst, M. Genotype constellation and evolution of group A rotaviruses infecting humans. *Curr. Opin. Virol.* **2**, 426–433 (2012).
15. McDonald, S. M., Davis, K., McAllen, J. K., Spiro, D. J. & Patton, J. T. Intra-genotypic diversity of archival G4P[8] human rotaviruses from Washington, DC. *Infect. Genet. Evol.* **11**, 1586–1594 (2011).
16. McDonald, S. M. *et al.* Evolutionary dynamics of human rotaviruses: balancing reassortment with preferred genome constellations. *PLoS. Pathog.* **5**, e1000634 (2009).
17. McDonald, S. M. *et al.* Diversity and relationships of cocirculating modern human rotaviruses revealed using large-scale comparative genomics. *J. Virol.* **86**, 9148–9162 (2012).
18. Silva, F. D., Espinoza, L. R., Tonietti, P. O., Barbosa, B. R. & Gregori, F. Whole-genomic analysis of 12 porcine group A rotaviruses isolated from symptomatic piglets in Brazil during the years of 2012-2013. *Infect. Genet. Evol.* **32**, 239–254 (2015).
19. Silva, F. D., Gregori, F. & McDonald, S. M. Distinguishing the genotype 1 genes and proteins of human Wa-like rotaviruses vs. porcine rotaviruses. *Infect. Genet. Evol.* **43**, 6–14 (2016).
20. Zeller, M. *et al.* Genome-Wide Evolutionary analyses of G1P[8] strains isolated before and after rotavirus vaccine introduction. *Genome Biol. Evol.* **7**, 2473–2483 (2015).
21. Zhang, S. *et al.* Analysis of human rotaviruses from a single location over an 18-year time span suggests that protein coadaption influences gene constellations. *J. Virol.* **88**, 9842–9863 (2014).
22. Gombold, J. L. & Ramig, R. F. Analysis of reassortment of genome segments in mice mixedly infected with rotaviruses SA11 and RRV. *J. Virol.* **57**, 110–116 (1986).
23. Graham, A., Kudesia, G., Allen, A. M. & Desselberger, U. Reassortment of human rotavirus possessing genome rearrangements with bovine rotavirus: evidence for host cell selection. *J. Gen. Virol.* **68**(Pt 1), 115–122 (1987).
24. Midthun, K., Hoshino, Y., Kapikian, A. Z. & Chanock, R. M. Single gene substitution rotavirus reassortants containing the major neutralization protein (VP7) of human rotavirus serotype 4. *J. Clin. Microbiol.* **24**, 822–826 (1986).
25. Ramig, R. F. & Ward, R. L. Genomic segment reassortment in rotaviruses and other reoviridae. *Adv. Virus Res.* **39**, 163–207 (1991).
26. McDonald, S. M. & Patton, J. T. Assortment and packaging of the segmented rotavirus genome. *Trends. Microbiol.* **19**, 136–144 (2011).
27. Li, W. *et al.* Genomic analysis of codon, sequence and structural conservation with selective biochemical-structure mapping reveals highly conserved and dynamic structures in rotavirus RNAs with potential cis-acting functions. *Nucleic Acids Res.* **38**, 7718–7735 (2010).
28. Suzuki, Y. A candidate packaging signal of human rotavirus differentiating Wa-like and DS-1-like genomic constellations. *Microbiol. Immunol.* **59**, 567–571 (2015).
29. Trask, S. D., McDonald, S. M. & Patton, J. T. Structural insights into the coupling of virion assembly and rotavirus replication. *Nat. Rev. Microbiol.* **10**, 165–177 (2012).
30. Trask, S. D., Taraporewala, Z. F., Boehme, K. W., Dermody, T. S. & Patton, J. T. Dual selection mechanisms drive efficient single-gene reverse genetics for rotavirus. *Proc. Natl. Acad. Sci. USA* **107**, 18652–18657 (2010).

31. Navarro, A., Trask, S. D. & Patton, J. T. Generation of genetically stable recombinant rotaviruses containing novel genome rearrangements and heterologous sequences by reverse genetics. *J. Virol.* **87**, 6211–6220 (2013).
32. Taraporewala, Z. F., Schuck, P., Ramig, R. F., Silvestri, L. & Patton, J. T. Analysis of a temperature-sensitive mutant rotavirus indicates that NSP2 octamers are the functional form of the protein. *J. Virol.* **76**, 7082–7093 (2002).
33. Jayaram, H., Taraporewala, Z., Patton, J. T. & Prasad, B. V. Rotavirus protein involved in genome replication and packaging exhibits a HIT-like fold. *Nature* **417**, 311–315 (2002).
34. Arnoldi, F., Campagna, M., Eichwald, C., Desselberger, U. & Burrone, O. R. Interaction of rotavirus polymerase VP1 with nonstructural protein NSP5 is stronger than that with NSP2. *J. Virol.* **81**, 2128–2137 (2007).
35. Eichwald, C., Rodriguez, J. F. & Burrone, O. R. Characterization of rotavirus NSP2/NSP5 interactions and the dynamics of viroplasm formation. *J. Gen. Virol.* **85**, 625–634 (2004).
36. Fabbretti, E., Afrikanova, I., Vascotto, F. & Burrone, O. R. Two non-structural rotavirus proteins, NSP2 and NSP5, form viroplasm-like structures *in vivo*. *J. Gen. Virol.* **80**(Pt 2), 333–339 (1999).
37. Hu, L. *et al*. Crystallographic analysis of rotavirus NSP2-RNA complex reveals specific recognition of 5′ GG sequence for RTPase activity. *J. Virol.* **86**, 10547–10557 (2012).
38. Taraporewala, Z. F. *et al*. Structure-function analysis of rotavirus NSP2 octamer by using a novel complementation system. *J. Virol.* **80**, 7984–7994 (2006).
39. Viskovska, M. *et al*. Probing the sites of interactions of rotaviral proteins involved in replication. *J. Virol.* **88**, 12866–12881 (2014).
40. Kanai, Y. *et al*. Entirely plasmid-based reverse genetics system for rotaviruses. *Proc. Natl. Acad. Sci. USA* **114**, 2349–2354 (2017).
41. Arnold, M., Patton, J. T. & McDonald, S. M. Culturing, storage, and quantification of rotaviruses. *Curr. Protoc. Microbiol.* Chapter 15, Unit 15C 13 (2009).
42. Ramig, R. F. Isolation and genetic characterization of temperature-sensitive mutants of simian rotavirus SA11. *Virology* **120**, 93–105 (1982).
43. Hofacker, I. L. Vienna RNA secondary structure server. *Nucleic Acids Res.* **31**, 3429–3431 (2003).
44. McDonald, S. M. & Patton, J. T. Rotavirus VP2 core shell regions critical for viral polymerase activation. *J. Virol.* **85**, 3095–3105 (2011).
45. McKell, A. O., LaConte, L. E., & McDonald, S. M. Temperature-sensitive lesion in the N-terminal domain of the rotavirus polymerase affects its intracellular localization and enzymatic activity. *J. Virol.* **13**, 91(7). pii: e00062-17. doi:10.1128/JVI.00062-17 (2017).
46. Pirolli, D. *et al*. Insights from molecular dynamics simulations: structural basis for the V567D mutation-induced instability of zebrafish alpha-dystroglycan and comparison with the murine model. *PLoS One* **9**, e103866 (2014).
47. Abraham, M. J. *et al*. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1**, 19–25 (2015).
48. Pettersen, E. F. *et al*. UCSF Chimera–a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
49. Lindorff-Larsen, K. *et al*. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **78**, 1950–1958 (2010).
50. Meinhold, L. & Smith, J. C. Fluctuations and correlations in crystalline protein dynamics: a simulation analysis of staphylococcal nuclease. *Biophys. J.* **88**, 255–2563 (2005).
51. Rueda, M. *et al*. A consensus view of protein dynamics. *Proc. Natl. Acad. Sci. USA* **104**, 769–801 (2007).

Acknowledgements

The authors would like to thank members of the McDonald laboratory for intellectual and technical support. We also thank Dr. John Patton (University of Maryland, College Park) for the generous donation of reagents. This work was supported through start-up funding from the Virginia Tech Carilion Research Institute and through grants from the National Institutes of Health (R01-AI116815 and R21-AI119588). C.P.L. was also supported by the Translational Biology, Medicine, and Health Graduate Program at Virginia Tech.

Author Contributions

S.M.M. designed the study, contributed to data analysis, and wrote the manuscript. R.M. and S.Z. performed all work related to the rescue and analysis of recombinant reassortant viruses and contributed equally. C.P.L. performed the *in vitro* analysis of NSP2 + RNA replication templates and generated homology models of NSP2. L.E.W.L. performed the molecular dynamics simulations of NSP2. All authors reviewed and edited the manuscript.

Additional Information

Supplementary information accompanies this paper at doi:10.1038/s41598-017-08068-w

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.