HHS Public Access

Author manuscript

Virology. Author manuscript; available in PMC 2017 January 01.

Published in final edited form as:

Virology. 2016 January ; 487: 150–162. doi:10.1016/j.virol.2015.10.017.

Modeling of the rotavirus group C capsid predicts a surface topology distinct from other rotavirus species

Elif Eren[1,*], **Kimberly Zamuda**[1,2], and **John T. Patton**[1,3]

[1]Laboratory of Infectious Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892

Abstract

Rotavirus C (RVC) causes sporadic gastroenteritis in adults and is an established enteric pathogen of swine. Because RVC strains grow poorly in cell culture, which hinders generation of virion-derived RVC triple-layered-particle (TLP) structures, we used the known Rotavirus A (RVA) capsid structure to model the human RVC (Bristol) capsid. Comparative analysis of RVA and RVC capsid proteins showed major differences at the VP7 layer, an important target region for vaccine development due to its antigenic properties. Our model predicted the presence of a surface extended loop in RVC, which could form a major antigenic site on the capsid. We analyzed variations in the glycosylation patterns among RV capsids and identified group specific conserved sites. In addition, our results showed a smaller RVC VP4 foot, which protrudes toward the intermediate VP6 layer, in comparison to that of RVA. Finally, our results showed major structural differences at the VP8* glycan recognition sites.

Keywords

rotavirus; capsid structure; antigenic topology

INTRODUCTION

Rotaviruses, members of the *Reoviridae*, are a common cause of severe gastroenteritis in humans and animals (1). The rotavirus genome consists of 11 segments of double-stranded RNA (dsRNA) that encode six structural proteins (VP1-VP4, VP6 and VP7) and five or six non-structural proteins (NSP1-NSP6) depending on the virus strain (1, 2). Rotaviruses have been resolved into 8 species (RVA to RVH) based on phylogenetic analysis (3). RVA, RVB, RVC and RVH have been recovered from humans and animal species, while RVD, RVF and RVG have been isolated only from birds and RVE from swine (4, 5). The RVA group is the most widely studied as its members are the leading cause of life-threatening diarrhea in

*Corresponding author: Elif Eren, Laboratory of Infectious Diseases, NIAID, NIH, 50 South Drive, Room 6312, Bethesda, MD 20892, USA. Phone: (240) 669-5303. elif.eren@nih.gov.

[2]College of William and Mary, Williamsburg, VA, 23186

[3]Virginia-Maryland College of Veterinary Medicine, University of Maryland, College Park, MD 20742

infants and young children globally. In contrast to the endemic incidence of RVA disease, RVC infections are associated with sporadic and limited outbreaks of gastroenteritis in children and adults. Such infections have been noted both in developed and developing countries (6–9). Many RVC infections may be asymptomatic or subclinical in nature, based on seroprevalence studies indicating that up to 50% of adults have anti-RVC antibodies in some geographic regions. RVC are also known to infect swine, cattle, dogs and ferrets and are a particularly important cause of gastroenteritis in pigs during the first 8 weeks of life. The widespread nature of swine RVC infections is suggested by seroprevalence studies revealing that 70 to 100% of adult pigs have anti-RVC antibodies.

High-resolution structures are available for the RVA capsid and most of its protein components. Based on these data, the RV virion has been determined to be a non-enveloped 100-nm icosahedral particle consisting of three concentric layers of protein (triple-layered particle (TLP) (1, 10–13). The innermost layer - the core shell - is a T=1 icosahedron composed of 60 asymmetric dimers of VP2 (VP2A and VP2B) that encloses the genome and two minor proteins: the RNA-dependent RNA polymerase VP1 and the RNA capping enzyme VP3. The intermediate layer has T=13 symmetry and is formed by 260 trimers of VP6. The outermost layer is composed of 260 trimers of the glycoprotein VP7, also arranged with T=13 symmetry, and 60 spikes formed by asymmetric VP4 trimers. Exposure to trypsin-like proteases cleaves RVA VP4 into two polypeptides, VP8* and VP5* and enhances virus infectivity (14). The VP7 and VP4 proteins contain multiple antigenic epitopes that can induce the production of neutralizing antibodies and, thus, are key targets for vaccine development.

A classification system has been adopted that allows assignment of G (glycoprotein) and P (protease-sensitive) genotypes to the VP7 and VP4 genes of RVA strains, respectively (1). To date, a total of 27 G and 37 P genotypes have been identified for the RVA. In contrast to RVA, much less genetic diversity has been found for the RVC; however, this may in part reflect the relatively fewer G and P-typing and surveillance studies that have been performed for the RVC. Although a VP7- and VP4- genotype classification system has yet to be formally established for the RVC, 9 G-genotypes (G1–G9) and 6 P-genotypes (P[1]–P[3]) have been proposed based on phylogenetic analysis of human and animal RVC sequences (15–19). The porcine RVC are the most diverse and include viruses with G1, G3, G5–G9P [1] genotypes. Typically, human RVCs have a G4P[2] genotype and bovine RVCs have a G2P[3] genotype (15, 17, 18, 20). There are reports suggesting interspecies transmission of RVC strains, but little information is available concerning the frequency of such events or its impact on RVC evolution or diversity. Dual infections by RVA and RVC in humans and animals have been reported (20). However, reassortment of RVA and RVC strains to form RVA/RVC hybrid viruses has not been described.

There is no structural information available for the RVC virion, including its outer capsid VP7 and VP4 components. The lack of such information hampers the ability to define antigenic regions of the RVC capsid that could be useful for vaccine development and monitoring virus diversity and evolution. Moreover, comparison of a RVC capsid structure with that of RVA would allow better understanding of virus features that are shared, or unique, among RV groups. RVC grows poorly in cell cultures, which makes it difficult to

obtain the necessary quantities of virus particles required for generating structure information by cryoelectron microscopy (Cryo-EM) or X-ray crystallography. Importantly, previous studies have shown that co-expression of the RVA intermediate capsid protein VP6 (A-VP6) with the RVC core protein VP2 (C-VP2) using recombinant baculoviruses generates C-VP2/A-VP6 double-layered particles (DLPs) (21, 22). Similarly, recombinant baculoviruses were also successfully used to produce RVA/RVC hybrid TLPs, including those composed of A-VP2/C-VP6/C-VP7, A-VP2/C-VP6/A-VP7 and A-VP2/A-VP6/C-VP7 (21). The ability to produce RVA/RVC hybrid particles indicates a high degree of structural conservation between RVC and RVA capsid proteins, and the presence of protein-protein interfaces that function similarly. Supported by these observations, we used a homology modeling based approach to generate a RVC capsid structure model, one found to score with a high confidence value. Notably, the model suggests that the surface topology of RVC capsid is different from that of the RVA capsid, principally due to an additional extended surface loop on the C-VP7 trimer that is likely to represent a dominant antigenic epitope. The C-VP4 attachment protein is also unique, containing a smaller VP6-interacting foot.

METHODS

Homology modeling and capsid generation

Human RVC (Bristol strain) VP2 (GenBank YP_392489), VP4 (YP_392514), and VP7 (YP_392513) were threaded against simian RVA (RRV strain) VP2, VP4, and VP7 (PDB IDs 3N09 and 3IYU) and modeled using the Phyre2 protein-fold recognition server (23). Similarly, human RVC (Bristol strain) VP6 (YP_392512) was threaded against bovine RVA (RF strain) VP6 (PDB ID 1QHD) and modeled by Phyre2. The models were evaluated using the protein structure validation software suite (PSVS) (http://psvs-1_5-dev.nesg.org/) (24). Model and template structure comparisons and Ca backbone root-mean-square deviation (RMSD) value calculations were carried out using PDBeFold (http://www.ebi.ac.uk/msd-srv/ssm) (25). The RVC 31-mer asymmetric unit was generated by structural alignment of RVC models with RVA asymmetric unit proteins (PDB IDs 3N09 and 3IYU) using Chimera (26) and by combination of aligned RVC structures to form the entire 31-mer asymmetric unit using PyMol (The PyMol Molecular Graphics System, Version 1.5.0.4., Schrodinger LLC). Minimization of the asymmetric unit was carried out using Chimera, applying 100 steepest descent steps and 3 conjugate gradient steps. Asymmetric unit protein geometries were evaluated using MolProbity (http://molprobity.biochem.duke.edu/) (27). A full capsid structure was modeled using Chimera after application of icosahedral capsid BIOMT values obtained from VIPERdb (http://viperdb.scripps.edu/) to the RVC asymmetric unit (28). Porcine G6 VP7 (ABR28507) was modeled using Phyre2, using RVA VP7 (PDB ID 3IYU) as a template. Model evaluation was carried out using PSVS and PDBeFold.

Sequence alignment and phylogenetic tree building

RVC VP4 and VP7 gene sequences were obtained by running BLAST analyses against human RVC (Bristol strain) VP4 and VP7 with the accession numbers NC_007572 and NC_007571, respectively. Best models for phylogenetic trees were analyzed by MEGA (version 6.06) (29). For both VP4 and VP7, Maximum Likelihood (ML) trees were constructed using general time reversible (GTR) model and G+I (gamma distributed with

invariant sites). RVC VP4 and VP7 protein sequences were obtained by running BLAST analyses against human RVC (Bristol strain) VP4 and VP7 with the accession numbers YP_392514 and YP_392513, respectively. Phylogenetic analyses and tree constructions of RVC VP4 and VP7 proteins were done by MEGA using the Neighbor-Joining method. Bootstrap analyses were carried out by MEGA using a value of 1000. Phylogenetic trees were edited using FigTree (v1.4.1). RVA (RRV strain) VP4 (ACC94315) and VP7 (ACC94320) sequences were obtained from the GenBank (30). Sequence alignments of RVA and RVC proteins were carried out using Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/) (31). RVA VP7 protein sequences were obtained by running a BLAST analysis against simian RVA (RRV strain) VP7 (ACC94320). RVA VP7 phylogenetic tree was constructed by MEGA using the Neighbor-Joining method. Sequence logo for the RVC VP7 insertion was created by WebLogo (http://weblogo.berkeley.edu/) (32).

Prediction of antigenic epitopes and glycosylation and trypsin-cleavage sites

Antigenic epitopes were predicted using EPCES (http://sysbio.unl.edu/EPCES/) (33) and structures were colored according to the antigenicity prediction scoring using PyMol, by application of B factor values generated by the program. N-linked glycosylation sites of RVC and RVA VP7 proteins were predicted using N-GlycoSite (http://www.hiv.lanl.gov/content/sequence/GLYCOSITE/glycosite.html) (34). Analysis of N-linked glycosylation site accessibility in VP7 structures was carried out using GlyProt (http://www.glycosciences.de/modeling/glyprot/php/main.php) (35). Trypsin site predictions were done using PeptideCutter (http://web.expasy.org/peptide_cutter/) (36).

VP8* structural alignments and ligand docking

Alignment of human RVC (Bristol) VP8* model with simian RVA (RRV) (PDB ID 1KQR) was carried out using DaliLite (http://www.ebi.ac.uk/Tools/structure/dalilite/) (37). Electrostatic surface potentials were calculated using PyMol APBS (Adaptive Poisson-Boltzmann Solver) tool.

RESULTS

Homology model of the human RVC capsid

High-resolution structural information has been obtained for the RVA TLP and its protein components through X-ray crystallography, NMR and single particle cryo-EM (10–13, 38–40). To produce a model of the RVC capsid VP2, VP4, VP6 and VP7 proteins of the human RVC (Bristol strain) were individually modeled using three different homology-modeling programs, Phyre2, I-TASSER and MODELLER, either by threading against a database of structures or user-specified RVA capsid protein structures. Phyre2 uses the alignment of hidden Markov models via HHsearch to significantly improve accuracy of alignment and detection rate (23). I-TASSER builds models based on multiple threading alignments by LOMETS (Local Meta-Threading Server) and iterative TASSER (Threading/Assembly/Refinement) simulations (41). MODELLER implements a technique known as *satisfaction of spatial restraints* that uses a set of geometrical criteria to create a probability density function for the location of each atom in the protein (42). In addition, MODELLER incorporates ab initio structure of protein loops, which are often highly variable and

therefore difficult to predict by homology modeling. The resultant models generated by the different homology modeling programs were evaluated using PSVS (protein structure validation software suite) that integrates analyses from several structure evaluation tools including PROCHECK, MolProbity, Verify3D and ProsaII, and PDBeFold that performs the multiple comparison and 3D alignment of protein structures. In general, I-TASSER and MODELLER generated models had slightly better Ramachandran-based restraints (1–3%) and lower bond angle (0.1–1.6Å) and bond length deviations (0.01–0.05Å) (data not shown). On the other hand, Phyre2 models yielded lower molprobity clashscores (6%–30%) and higher overall model qualities (0–40%) (data not shown). In addition, when compared to the template structures Phyre2 models yielded lowest root mean square deviation (RMSD) values and highest Q-scores (quality of alignment), which represent the quality function of Cα alignment, maximized by the SSM (secondary structure matching) algorithm (data not shown). Based on these assessments, Phyre2 models of RVC VP2, VP4 and VP7 threaded against the corresponding capsid proteins of the RVA RRV strain (PDB IDs 3N09 and 3IYU) and RVC VP6 threaded against the RVA RF strain (PDB ID 1QHD) were chosen to generate the RVC capsid model (13, 40).

Alignment of RVC capsid protein models with RVA capsid proteins revealed a high sequence similarity between RVC/RVA VP2 and VP6 proteins (>43% identity), while RVC/RVA VP7 and VP4 were more divergent (32–35% [Table 1]). All RVC models yielded favorable Ramachandran-based restraints (>93% in allowed regions) and comparison of model structures with the template structures resulted in low RMSD values (<0.6) for the Cα backbone (Table 1). Accordingly, Q-scores were >0.8 (1 is the highest score) indicating a good structural alignment between the models and the templates. All models yielded above average overall quality as assessed by Verify3D which scores the compatibility of the 3D structure model with its amino acid sequence (Table 1) (43). In order to generate a RVC 31-mer asymmetric unit, individual RVC models were aligned with corresponding RVA asymmetric unit proteins using Chimera (Fig. 1A). Protein geometries of both the RVC (model) and RVA (template) asymmetric units were evaluated using MolProbity. RVC model yielded a comparable MolProbity score in comparison to RVA template (3.93 vs. 4.45). MolProbity score combines the clashscore, rotamer and Ramachandran evaluations into a single score and a lower score is an indication of “better” geometry. Full RVC capsid model was produced from the asymmetric unit using Chimera after application of BIOMT values for icosahedral capsids (Fig. 1B). RVC capsid model was deposited to VIPERdb with the accession numbers HRTVC1 (VP2-VP6 layer) and HRTVC2 (VP7-VP4 layer) (http://viperdb.scripps.edu/info_page.php?VDB=hrtvc1 and http://viperdb.scripps.edu/info_page.php?VDB=hrtvc2).

Full RVC capsid model was composed of 60 asymmetric VP2 dimers (A and B forms) arranged in T=1 symmetry, 260 VP6 and VP7 homotrimers arranged in T=13 symmetry and 60 VP4 heterotrimers. The VP2 layer lacked the N-terminus aas 1–103 for VP2A and 1–85 for VP2B. The N-terminus of VP2B forms flexible tethers (aa 1–92) that are thought to interact with VP1 and VP3, and play an important role in their encapsidation (11, 44). In the RVC capsid model VP4 was modeled to mimic the unusual symmetry of the RVA VP4 spike. Accordingly, the base of the spike formed a trimer while the VP5* β-barrel domains and the VP8* heads of only two VP4 molecules formed a dimer. VP5* β-barrel domain of

the third VP4 molecule formed a diagonal “stalk” and the third VP8* head was missing from the structure model.

Comparison of the RVA capsid and RVC capsid model showed that both strains formed similarly sized and shaped DLPs and TLPs. However, RVC capsid model contained “peak” regions at the VP7 layer that are not observed in RVA capsid structure (Fig. 1C). Therefore, we analyzed RVC VP7 proteins and compared them to RVA VP7 proteins.

RVC capsid glycoprotein VP7

In a previous study, 70 RVC VP7 proteins were resolved into a number of distinct genotypes that displayed differences in host specificities (18). Since the number of available RVC VP7 sequences increased considerably, we reanalyzed the phylogenetic variability of 168 RVC VP7 sequences from human, porcine and bovine species, using the same 85% nucleotide cut-off value (Fig. 2A). Our results were in agreement with the previous study that defined 9 distinct G-types, 7 of which were formed by porcine strains. In order to compare the phenotypic variation between RVC VP7s from different host groups we analyzed RVC VP7 protein sequences (N=168). Phylogenetic analysis of amino acid sequences revealed that both human and bovine RVC VP7 proteins were highly conserved (>95% sequence identity) with each clustering into single groups (Fig. 2B and Table 2). On the other hand, porcine RVC VP7 proteins were divergent (69–99.7% sequence identity), forming 4 major distinct groups (G3, G5/G9, G1/G7 and G6/G8) with an 87% amino acid cut-off (Fig. 2B and Table 2). Among these porcine-G3 VP7 shared the lowest intergroup similarity with the other porcine groups (68.7–75.3%) (Table 2).

Sequence alignment of RVC VP7s with RVA VP7 showed the presence a 9 amino acid (aa) long insertion (L9) located between the RVC VP7 amino acids (aa) 83 and 93 (Fig. 3A). L9 is mostly composed of non-polar amino acids and is present in all RVC VP7s with some variations in the amino acid composition (Fig. 3B). The most common sequences are A[VA][VI]GS[QP]GPG which are found in over 80% of RVC VP7 proteins (Fig. 3B). Interestingly, most porcine-G6 RVC VP7s (>94%) were longer than the other RVC VP7s (736 aa vs. 732 aa) and had an extra 4 aa serine-rich SS[ST][VLM] insertion (L4a) at the aa 244–245 (numbering is relative to human RVC) position (Fig. 3A). L4a immediately precedes a 4 aa variable surface loop ($NX_1X_2X_3$ [L4b]) (aa 245–248) found in both RVA and RVC VP7s (Fig. 3A). In order to determine the positions of these insertions in the VP7 structure we analyzed human and porcine-G6 RVC VP7 models and compared them to the RVA VP7 structure. In both the human and porcine-G6 VP7 models L9 forms a surface exposed loop (Fig. 4A) that matches the VP7 surface peaks in the RVC capsid model. In porcine-G6 VP7 model, L4a and L4b together form an elongated surface exposed loop (L4) (Fig. 4A, bottom).

Impact of RVC VP7 model L9 and L4a loops on surface topology

The positions of antigenic regions on VP7 trimer surfaces for RVA have been determined by identification of neutralization escape mutants using various monoclonal antibodies (mABs) (38). RVA antigenic regions cluster roughly into two regions designated 7-1 and 7-2 (Fig. 4B, left). Region 7-1 spans the intersubunit boundary, and further divided into two

subregions: 7-1a, on one side of the interface, and 7-1b on the other (Fig. 4B, left). Among these 7-1 is the “immunodominant” region that is recognized by 58 out of 68 tested antibodies (38). We mapped the RVA antigenic regions on human and porcine-G6 RVC VP7 models to compare the topology of antigenic regions on VP7 surfaces and analyzed the effects of L9 and L4a insertions on these regions. Since RVC/RVA VP7 backbones were highly similar (RMSD 0.11 and Q-score 0.96) most RVA antigenic regions were located in roughly the same positions on RVC VP7 model surfaces (Fig. 4B). However, L9 insertion resulted in an altered RVC VP7 model surface topology at the 7-1a region (Fig. 4B, middle and right). On the other hand, in the porcine-G6 VP7 model L4a was inserted into the 7-2 region (Fig. 4B, right).

In order to determine the antigenic properties of L9 and L4, we analyzed both human and porcine-G6 VP7 model structures using EPCES (Epitope Prediction by Consensus Scoring). EPCES uses six different scoring functions including residue epitope propensity, conservation score, side-chain energy score, contact number, surface planarity score and secondary structure composition. Not surprisingly, both loops received high antigenicity scores (>90%) that matched with their surface location (Fig. 5).

RVC VP7 N-linked glycosylation sites

It has been shown that N-linked glycosylation of RV VP7 is required for correct disulfide bond formation and folding of the protein in the endoplasmic reticulum (ER) (45). We analyzed potential N-linked glycosylation sites of both RVA and RVC VP7s and found that there were three mainly conserved N-linked glycosylation sites in each group. These sites corresponded to aas 69, 146 and 238 in RVA, and 67, 152, and 225 in RVC (Fig. 6A, left and middle). Our results showed that RVs could have different degrees of N-linked glycosylation. Around 34% of RVA VP7s had a single glycosylation site at position 69 and 46% had two sites at positions 69 and 238. A 3.3% population corresponding to bovine/lamb RVA G10-types had two sites at positions 69 and 145. Interestingly, only 13.3% RVA VP7s had all three sites. The remaining 3.4% were more divergent and had 1–4 sites mostly at non-conserved positions. When we analyzed RVA VP7s with three sites we found that most of these clustered in a single group of human RVA G2-type VP7s (Fig. 6B). Most RVC VP7s had either two sites (porcine groups) at positions 67 and 225, or three sites at positions 67, 152 and 225 (human and porcine) or positions 67, 225 and 270 (bovine) (Fig. 6A, right). Interestingly, almost all human and bovine RVC VP7s had three sites at the respective conserved positions. In contrast, only a small population of porcine VP7s (17%) had all three sites and a small percentage (< 3%) had 2–3 sites at non-conserved positions.

Structural alignment of simian RVA VP7 and human RVC VP7 model showed that conserved N-linked glycosylation sites were not aligned but in close proximity for positions 67/69 and 146/152 (Fig. 6C). On the contrary, the asparagine residues 225/238 were located in distant positions on the opposite sites of the VP7 surface indicating that this location was group specific (Fig. 6C). Of the conserved sites only 67/69 are located on the N-terminus and not on the protein surface. Glycosylation at position 69 for RVA (SA11 strain) has been confirmed by biochemical assays (45, 46). In addition, in the RVA capsid structure a density is visible in close proximity to N69 that has been modeled as N-acetyl-D-glucosamine

molecules (NAG) (13). Studies related to the characterization of oligosaccharide compositions of RVs has shown that oligosaccharides of both simian (SA11) RVA and calf RVA likely exist as multiple high-mannose forms (46, 48). Glycosylation of surface located conserved sites has not been shown yet. However, it has been shown that a mutation at position 211 of RVA that creates a new N-linked glycosylation site prevents the antibody recognition of region 7-1, indirectly showing that surface N-linked glycosylation in RVA is possible (47).

RVC attachment protein VP4

RV VP4 proteins form spikes that project from the outer layer of mature virions (Fig 1A and 1B). VP4 mediates virion attachment to the host epithelial cell receptors and plays a major role in cell penetration, hemagglutination and virulence (49–53). VP4 is susceptible to proteolysis and it is cleaved by trypsin into two domains, VP8* and VP5* (Fig. 1A) (54–57). Trypsin cleavage of VP4 appears to increase RV infectivity, probably by ordering otherwise flexible uncleaved spikes (13, 55). VP8* binds to host glycans in order to facilitate host cell entry (58–60) and in integrin-dependent strains VP5* is thought to target the integrin heterodimer ITGA2/ITGB1 for cell attachment (61, 62). In VP4, VP8* is the major target for neutralizing antibodies.

A recent study that involved the phylogenetic analysis of 22 porcine VP4s revealed the presence of 6 clusters (19). We extended that analysis to 70 human, porcine and bovine RVC VP4 genes using an 78% nucleotide cut-off and identified 9 P-types some of which were divided into subgroups (as shown in roman numerals) based on the length of their nucleotide sequence in accordance to the system used by Suzuki et al., 2015 (Fig. 7A). Similar to VP7, human and bovine VP4s clustered in single groups while porcine VP4s showed higher variability (7 P-types), which was also observed for VP4 proteins (Fig. 7A and B). Our amino acid sequence analysis also showed that compared to RVA VP4s, which show a high degree of sequence variability (53–99.6% sequence identity), RVC VP4s were more conserved (68–99.2% sequence identity) (Table 3).

Human RVC VP4 proteins were longer in comparison to the other RVC VP4s (744 aa). Bovine VP4s varied between 733–735 aa while porcine VP4s varied between 729–736 aa. Therefore, we compared RVC VP4s and analyzed the positions of insertions and deletions with respect to each other (Fig. 8A). Interestingly, the majority of differences involve single or double aa insertions or deletions (indels) mapped to VP8*, which plays an important role in host receptor recognition and contains major antigenic epitopes (Fig. 8B, left). The biggest structural difference between RVC VP8* was caused by a 4 aa deletion between aa 209–213 which was observed in some bovine and porcine strains (Fig. 8B, left). This position partially involves β12 and the corresponding loop that connects it to β13 in the human VP8* structure. Another difference between human RVC VP8* and other RVC VP8* was an insertion between aa 251–256 that corresponded to the foot region of VP5* (Fig. 8B, right). However, due to trypsin cleavage of RVA VP5* that region could only be seen in one of the VP4 molecules of the heterotrimer, interacting with the rest of the foot region after cleavage.

In human RVC (Bristol) VP4 we identified six potential trypsin cleavage sites (K231, K244, K247, K248, K254 and K268) (Fig. 9A). Of these, [KR]231 was conserved in all RVC VP4 proteins and K268 (K258 for RVA) was conserved in all RVC and RVA VP4s. It has been shown that RVA VP4 is cleaved at K258 followed by the initial cleavage of VP4 at positions 231, 241 and 247 (55, 63).

Sequence alignment of human RVC (Bristol) VP4 model and RVA (RRV) VP4 showed that RVA VP4 had a unique serine/threonine/arginine rich insertion (STR-insertion, aa 571–620) located at the "foot" of VP5* (Fig. 9A and 9B). This insertion is found in all RVA VP4s with a high degree of conservation of serine and threonine residues. However, to date no phosphorylation has been reported for RVA VP4, and the role of the STR-insertion, if any, has yet to be determined. Since in the capsid VP5* interacts both with the VP7 and VP6 layers it is not clear whether the absence of this loop affects the backbone structures of these layers in the RVC capsid. VP4 proteins are held in the capsid by multiple points of interactions with VP6 (13). In addition, production of TLPs composed of RVA/RVC hybrid VP6 and VP7 layers suggest a high similarity between the backbone structures of these layers in both capsids (21). We hypothesized that the absence of the STR-insertion probably would not induce large conformational changes in the backbone structures of RVC capsid proteins. Analysis of RVA capsid structure revealed that the STR-insertion extended towards the VP7 N-terminus and one of the VP6 loops (Fig. 9C). However, the lack of high-resolution density at the foot region of RVA prevented the analysis of protein-protein interactions in detail so that we could not speculate about a direct interaction of this region with either the VP7 N-terminus or VP6, affecting the locking of VP7 layer in the RVA capsid. In RVC capsid model, reduced VP4 foot at that region is not in close proximity to either VP7 N-terminus or VP6 to form any direct hydrogen bonding interactions.

RVC VP4 glycan binding sites

RVA VP8* proteins can be grouped into sialidase-sensitive and sialidase-insensitive strains depending on their requirement to bind to a host neuraminidase-sensitive sialic acid (SA) moiety to enter the cell (50, 64). Most RVA strains, including a large majority of human RVAs, are sialidase-insensitive (58). Sialidase-sensitive animal RVAs bind to glycans with a terminal SA, whereas sialidase-insensitive human and animal strains bind to glycans with either an internal SA, such as GM1 (65) or other non sialylated-glycans such as Atype histo-blood group antigen (HBGA) (39). Although, the SA binding properties of RVC VP8* proteins have not been extensively characterized, one report suggested that a porcine RVC VP8* (strain AmC-1) was sialidase-sensitive (66). In order gain insight into the glycan recognition by RVC VP4 proteins, we compared RVC (Bristol) VP8* structure model to the sialidase-sensitive simian RVA (RRV [PDB ID 1KQR]) structure that contains a 2-O-methyl-5-N-acetyl-alpha-D-neuraminic acid (MNA) (39, 67). Similar to RVA VP8* proteins, RVC VP8* model had the beta-sandwich fold of galectins, a family of sugar binding proteins (68–70). In RVA sialidase-sensitive strains, the key SA binding residues strongly conserved (67). Among these, the conservation of R101, Y189, and S190, is absolute, while residues Y155 and Y189 can be mutated to histidines (Fig. 10A, left) (67). In Simian RVA VP8* structure, SA binds in a shallow groove between its two β-sheets (67). The groove has a large electropositive surface formed mainly by R101 that coordinates the

oxygen atoms of the glycerol group of the SA (Fig. 10B, left). The SA acetamido amino group points towards an electronegative surface formed by Y188 which forms a hydrogen bond with the acetamido nitrogen, and S190 hydrogen bonds with the SA carboxylate group (Fig. 10B, left). Sequence and structural alignment of the RVA VP8* substrate binding site with RVC VP8* model showed that the amino acid composition of substrate coordinating residues and surface charge distribution in these structures were quite different (Fig. 10A and 10B) suggesting that this site was probably not feasible for SA coordination in RVC. Supporting this, docking of MNA to this site did not provide satisfactory results. On the other hand, blind SA docking to RVC VP8* model using AutoDock Vina (71) resulted in identification of a number of potential SA binding sites with calculated binding affinities ranging from −4.7 kcal/mol to −5.4 kcal/mol (data not shown). However, that did not suggest an absolute requirement of a terminal SA for host cell entry.

It has been shown that RRV VP5* and human RVA (Wa strain) contains the α2β1 ligand sequence D308GE310 and D308 and G309 are necessary for the binding of VP5* to I domain of the α2 integrin subunit (72). We compared RVC VP5* and RVA VP5* sequences to analyze if RVC could also recognize α2β1. In RVC VP5* proteins the conserved DGE were replaced by conserved “[AS]GE” (Fig. 9A). Site-directed mutagenesis studies has shown that a double mutant (D308A/G309A) lost its ability to bind to α2β1 (72). Therefore, it is unlikely that RVC VP5* will recognize α2β1 by its [AS]GE sequence. However, RVC VP5* might contain other domains that facilitate integrin-independent cell entry as in the case of porcine RVA CRW-8 VP5* (72).

DISCUSSION

Virus capsid structures play an important role in vaccine development, drug targeting and improvement of immunodiagnostic tools. Although the most widely used methods for determination of capsid structures are cryo-EM and X-ray crystallography, the availability of new bioinformatics tools enables the generation of capsid models for virus strains that are hard to crystallize or cannot be obtained in high quantities necessary for these applications. One such example is the human rhinovirus C (C15) capsid model developed by Basta et al., which enabled the authors to determine the underlying reasons for the antiviral drug resistance of this strain (73).

To date at least 8 different RV species have been identified. However, structural information is only available for RVA, partly due to difficulties in the adaptation of other strains to growth in cell culture. To this end, we generated a RVC capsid model by homology modeling and compared RVA capsid and RVC capsid model in order to understand the structural features that might underlie the differences between the pathogenicity of these two RV groups. This comparison enabled us to define new group/host specific (L4) or generally conserved (L9) surface exposed loops at the RVC VP7 layer that potentially form major antigenic sites which are absent from the RVA capsid surface. L4 is found only in a specific group (G6) of porcine-RVCs and probably does not play an important role in VP7 physiology. On the other hand, L9 is present in all RVCs regardless of host specificity and its sequence has a high degree of conservation. Although, at this point it is not clear whether L9 forms a novel domain that plays a role in RVC pathogenicity i.e. host cell/compartment

entry, it can be used in seroepidemiological studies for RVC identification. In addition, its surface exposed location in VP7 makes it a good target for genetic engineering of RVC or RVA outer capsids *in vitro*.

Glycoproteins figure prominently in a number of viruses and mutations at the surface exposed glycosylation sites often interfere with viral entry, infectivity, tissue specificity, or host range, implying roles in cell recognition, membrane fusion and cell entry (74, 75). Our analysis showed that both RVA and RVC VP7 proteins have potential N-linked glycosylation sites at conserved locations at the N-terminus or on the surface. However, both strains could adopt different levels of glycosylation (0–4 sites for RVA and 2–3 sites for RVC) and in general RVC VP7s contained a higher number of potential N-linked glycosylation sites in comparison to RVA VP7s (1–2 sites for RVA and 2–3 sites for RVC). RVC VP7 model has an increased surface area with highly probable antigenic properties suggesting that there might be a correlation between the degree of accessible antigenic regions and the number of N-linked glycosylation sites. Indeed, as has been suggested for HIV, glycosylation sites provide variability that allows the virus to escape immune detection of nearby conserved amino acids at the receptor-binding site and modulate antigenicity (76, 77). Further supporting this, in RV it has been possible to select monoclonal neutralization escape mutants that have new glycosylation sites in a defined epitope (78).

RVA-RVC hybrid TLPs composed of VP2, VP6 and VP7 have been produced indicating that the reassortment of these viruses is possible at the capsid level. However, no research has been conducted to test the incorporation of VP4, which shows the highest degree of variability among capsid proteins, in these hybrid TLPs. Our homology model and sequence alignment showed that RVA and RVC VP4 differ at the foot region, which protrudes towards the intermediate capsid layer. Even though our model and hybrid TLPs suggest a similar backbone arrangement at the VP2, VP6, and VP7 layer for both RVA and RVC capsids, further structural and biochemical studies are required to determine whether VP4 presents a structural barrier for the reassortment of these viruses.

Comparison of RVA and RVC VP8* molecules showed no similarity between the two structures at the SA binding site of RVA. One possibility is the recognition of a different receptor by this site and another is the presence of alternative binding sites on the RVC VP8* surface. Previous biochemical studies showed a terminal SA residue dependence of a porcine RVC strain for host cell entry (66). We identified potential binding pockets on the RVC VP8* model surface that might interact with SA. However, these sites have yet to be verified either by structural or site-directed mutagenesis experiments. Also, it is possible that SA-independent RVC strains exist as in the case of RVA. Interestingly, RVCs lack the conserved sequences at the VP5* integrin binding sites underlying differences in host receptor recognition by RVA and RVC strains. We found insertion/deletion mutations among different RVC groups in the VP8*, where most antigenic regions of the RV VP4 are located. It is a common mechanism to generate and maintain mutations that allow viruses to escape from antibody neutralization, so further comparative analysis of these regions could lead to a better understanding of how RVs evolve under selective immune pressure.

In conclusion, our modeled RVC capsid provided a useful tool to study the biology of this virus group from a structural perspective and enabled us to identify some structural differences that probably underlie the differences between the antigenicity and host cell receptor recognition by these two strains. We believe our work provides a framework for comparative studies with other rotaviruses.

Acknowledgments

We would like to thank to the members of the laboratory for critical review of the manuscript and helpful discussions. This work was supported by the Intramural Research Program of the National Institute of Allergy and Infectious Diseases, National Institutes of Health.

References

1. Estes, MK.; Kapikian, A. Fileds Virology. 5. Williams and Wilkins; Philedelphia: 2007. Rotaviruses; p. 1917-1974.
2. Ciarlet, M.; Estes, MK. Encyclopedia of Environmental Microbiology. John Wiley & Sons; New York: 2002. Rotaviruses: basic biology, epidemiology and methodologies; p. 2573-2773.
3. Matthijnssens J, Otto PH, Ciarlet M, Desselberger U, Van Ranst M, Johne R. VP6-sequence-based cutoff values as a criterion for rotavirus species demarcation. Arch Virol. 2012; 157:1177–1182. [PubMed: 22430951]
4. Patton JT. Rotavirus diversity and evolution in the post-vaccine world. Discov Med. 2012; 13:85–97. [PubMed: 22284787]
5. Wakuda M, Ide T, Sasaki J, Komoto S, Ishii J, Sanekata T, Taniguchi K. Porcine rotavirus closely related to novel group of human rotaviruses. Emerg Infect Dis. 2011; 17:1491–1493. [PubMed: 21801631]
6. Kuzuya M, Fujii R, Hamano M, Yamada M, Shinozaki K, Sasagawa A, Hasegawa S, Kawamoto H, Matsumoto K, Kawamoto A, Itagaki A, Funatsumaru S, Urasawa S. Survey of human group C rotaviruses in Japan during the winter of 1992 to 1993. J Clin Microbiol. 1998; 36:6–10. [PubMed: 9431910]
7. Nilsson M, Svenungsson B, Hedlund KO, Uhnoo I, Lagergren A, Akre T, Svensson L. Incidence and genetic diversity of group C rotavirus among adults. J Infect Dis. 2000; 182:678–684. [PubMed: 10950759]
8. Rahman M, Banik S, Faruque AS, Taniguchi K, Sack DA, Van Ranst M, Azim T. Detection and characterization of human group C rotaviruses in Bangladesh. J Clin Microbiol. 2005; 43:4460–4465. [PubMed: 16145092]
9. Von Bonsdorf CH, Svensson L. Human serogroup C rotavirus in Finland. Scand J Infect Dis. 1988; 20:475–478. [PubMed: 2851876]
10. Chen JZ, Settembre EC, Aoki ST, Zhang X, Bellamy AR, Dormitzer PR, Harrison SC, Grigorieff N. Molecular interactions in rotavirus assembly and uncoating seen by high-resolution cryo-EM. Proc Natl Acad Sci U S A. 2009; 106:10644–10648. [PubMed: 19487668]
11. Estrozi LF, Settembre EC, Goret G, McClain B, Zhang X, Chen JZ, Grigorieff N, Harrison SC. Location of the dsRNA-dependent polymerase, VP1, in rotavirus particles. J Mol Biol. 2013; 425:124–132. [PubMed: 23089332]
12. McClain B, Settembre E, Temple BR, Bellamy AR, Harrison SC. X-ray crystal structure of the rotavirus inner capsid particle at 3.8 A resolution. J Mol Biol. 2010; 397:587–599. [PubMed: 20122940]
13. Settembre EC, Chen JZ, Dormitzer PR, Grigorieff N, Harrison SC. Atomic model of an infectious rotavirus particle. EMBO J. 2011; 30:408–416. [PubMed: 21157433]
14. Dormitzer PR, Nason EB, Prasad BV, Harrison SC. Structural rearrangements in the membrane penetration protein of a non-enveloped virus. Nature. 2004; 430:1053–1058. [PubMed: 15329727]
15. Collins PJ, Martella V, O'Shea H. Detection and characterization of group C rotaviruses in asymptomatic piglets in Ireland. J Clin Microbiol. 2008; 46:2973–2979. [PubMed: 18632912]

16. Jiang B, Gentsch JR, Tsunemitsu H, Saif LJ, Glass RI. Sequence analysis of the gene encoding VP4 of a bovine group C rotavirus: molecular evidence for a new P genotype. Virus Genes. 1999; 19:85–88. [PubMed: 10499454]
17. Martella V, Banyai K, Lorusso E, Decaro N, Bellacicco A, Desario C, Corrente M, Greco G, Moschidou P, Tempesta M, Arista S, Ciarlet M, Lavazza A, Buonavoglia C. Genetic heterogeneity in the VP7 of group C rotaviruses. Virology. 2007; 367:358–366. [PubMed: 17614111]
18. Marthaler D, Rossow K, Culhane M, Collins J, Goyal S, Ciarlet M, Matthijnssens J. Identification, phylogenetic analysis and classification of porcine group C rotavirus VP7 sequences from the United States and Canada. Virology. 2013; 446:189–198. [PubMed: 24074581]
19. Suzuki T, Hasebe A, Miyazaki A, Tsunemitsu H. Analysis of genetic divergence among strains of porcine rotavirus C, with focus on VP4 and VP7 genotypes in Japan. Virus Res. 2015; 197:26–34. [PubMed: 25499298]
20. Jiang B, Dennehy PH, Spangenberger S, Gentsch JR, Glass RI. First detection of group C rotavirus in fecal specimens of children with diarrhea in the United States. J Infect Dis. 1995; 172:45–50. [PubMed: 7797945]
21. Kim Y, Chang KO, Kim WY, Saif LJ. Production of hybrid double- or triple-layered virus-like particles of group A and C rotaviruses using a baculovirus expression system. Virology. 2002; 302:1–8. [PubMed: 12429511]
22. Tosser G, Labbe M, Bremont M, Cohen J. Expression of the major capsid protein VP6 of group C rotavirus and synthesis of chimeric single-shelled particles by using recombinant baculoviruses. Journal of virology. 1992; 66:5825–5831. [PubMed: 1326644]
23. Kelley LA, Sternberg MJ. Protein structure prediction on the Web: a case study using the Phyre server. Nat Protoc. 2009; 4:363–371. [PubMed: 19247286]
24. Bhattacharya A, Tejero R, Montelione GT. Evaluating protein structures determined by structural genomics consortia. Proteins. 2007; 66:778–795. [PubMed: 17186527]
25. Krissinel E, Henrick K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. Acta Crystallogr D Biol Crystallogr. 2004; 60:2256–2268. [PubMed: 15572779]
26. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem. 2004; 25:1605–1612. [PubMed: 15264254]
27. Chen VB, Arendall WB 3rd, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC. MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D Biol Crystallogr. 2010; 66:12–21. [PubMed: 20057044]
28. Shepherd CM, Borelli IA, Lander G, Natarajan P, Siddavanahalli V, Bajaj C, Johnson JE, Brooks CL 3rd, Reddy VS. VIPERdb: a relational database for structural virology. Nucleic Acids Res. 2006; 34:D386–389. [PubMed: 16381893]
29. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular evolutionary genetics analysis version 6.0. Mol Biol Evol. 2013; 30:2725–2729. [PubMed: 24132122]
30. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL. GenBank. Nucleic Acids Res. 2005; 33:D34–38. [PubMed: 15608212]
31. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Soding J, Thompson JD, Higgins DG. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol. 2011; 7:539. [PubMed: 21988835]
32. Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: A sequence logo generator. Genome Research. 2004; 14:1188–1190. [PubMed: 15173120]
33. Liang S, Liu S, Zhang C, Zhou Y. A simple reference state makes a significant improvement in near-native selections from structurally refined docking decoys. Proteins. 2007; 69:244–253. [PubMed: 17623864]
34. Shaw TI, Zhang M. HIV N-linked glycosylation site analyzer and its further usage in anchored alignment. Nucleic Acids Res. 2013; 41:W454–458. [PubMed: 23748959]
35. Bohne-Lang A, von der Lieth CW. GlyProt: in silico glycosylation of proteins. Nucleic acids research. 2005; 33:W214–219. [PubMed: 15980456]

36. Gasteiger, E.; Hoogland, C.; Gattiker, A.; Duvaud, SD.; Wilkins, MR.; Appel, RD.; Bairoch, A. Protein Identification and Analysis Tools on the ExPASy Server. Humana Press; 2005.
37. Hasegawa H, Holm L. Advances and pitfalls of protein structural alignment. Current opinion in structural biology. 2009; 19:341–348. [PubMed: 19481444]
38. Aoki ST, Settembre EC, Trask SD, Greenberg HB, Harrison SC, Dormitzer PR. Structure of rotavirus outer-layer protein VP7 bound with a neutralizing Fab. Science. 2009; 324:1444–1447. [PubMed: 19520960]
39. Hu L, Crawford SE, Czako R, Cortes-Penfield NW, Smith DF, Le Pendu J, Estes MK, Prasad BV. Cell attachment protein VP8* of a human rotavirus specifically interacts with A-type histo-blood group antigen. Nature. 2012; 485:256–259. [PubMed: 22504179]
40. Mathieu M, Petitpas I, Navaza J, Lepault J, Kohli E, Pothier P, Prasad BV, Cohen J, Rey FA. Atomic structure of the major capsid protein of rotavirus: implications for the architecture of the virion. EMBO J. 2001; 20:1485–1497. [PubMed: 11285213]
41. Zhang Y. I-TASSER server for protein 3D structure prediction. BMC Bioinformatics. 2008; 9:40. [PubMed: 18215316]
42. Fiser A, Sali A. Modeller: generation and refinement of homology-based protein structure models. Methods Enzymol. 2003; 374:461–491. [PubMed: 14696385]
43. Eisenberg D, Luthy R, Bowie JU. VERIFY3D: assessment of protein models with three-dimensional profiles. Methods Enzymol. 1997; 277:396–404. [PubMed: 9379925]
44. Zeng CQ, Estes MK, Charpilienne A, Cohen J. The N terminus of rotavirus VP2 is necessary for encapsidation of VP1 and VP3. J Virol. 1998; 72:201–208. [PubMed: 9420216]
45. Mirazimi A, Svensson L. Carbohydrates facilitate correct disulfide bond formation and folding of rotavirus VP7. J Virol. 1998; 72:3887–3892. [PubMed: 9557673]
46. Both GW, Siegman LJ, Bellamy AR, Atkinson PH. Coding assignment and nucleotide sequence of simian rotavirus SA11 gene segment 10: location of glycosylation sites suggests that the signal peptide is not cleaved. Journal of virology. 1983; 48:335–339. [PubMed: 6312090]
47. Mackow ER, Shaw RD, Matsui SM, Vo PT, Benfield DA, Greenberg HB. Characterization of homotypic and heterotypic VP7 neutralization sites of rhesus rotavirus. Virology. 1988; 165:511–517. [PubMed: 2457279]
48. Kouvelos K, Petric M, Middleton PJ. Oligosaccharide composition of calf rotavirus. The Journal of general virology. 1984; 65(Pt 7):1159–1164. [PubMed: 6086809]
49. Crawford SE, Labbe M, Cohen J, Burroughs MH, Zhou YJ, Estes MK. Characterization of virus-like particles produced by the expression of rotavirus capsid proteins in insect cells. J Virol. 1994; 68:5945–5952. [PubMed: 8057471]
50. Fiore L, Greenberg HB, Mackow ER. The VP8 fragment of VP4 is the rhesus rotavirus hemagglutinin. Virology. 1991; 181:553–563. [PubMed: 1849677]
51. Kirkwood CD, Bishop RF, Coulson BS. Attachment and growth of human rotaviruses RV-3 and S12/85 in Caco-2 cells depend on VP4. J Virol. 1998; 72:9348–9352. [PubMed: 9765487]
52. Ludert JE, Feng N, Yu JH, Broome RL, Hoshino Y, Greenberg HB. Genetic mapping indicates that VP4 is the rotavirus cell attachment protein in vitro and in vivo. J Virol. 1996; 70:487–493. [PubMed: 8523562]
53. Trask SD, McDonald SM, Patton JT. Structural insights into the coupling of virion assembly and rotavirus replication. Nat Rev Microbiol. 2012; 10:165–177. [PubMed: 22266782]
54. Clark SM, Roth JR, Clark ML, Barnett BB, Spendlove RS. Trypsin enhancement of rotavirus infectivity: mechanism of enhancement. J Virol. 1981; 39:816–822. [PubMed: 6169841]
55. Crawford SE, Mukherjee SK, Estes MK, Lawton JA, Shaw AL, Ramig RF, Prasad BV. Trypsin cleavage stabilizes the rotavirus VP4 spike. J Virol. 2001; 75:6052–6061. [PubMed: 11390607]
56. Espejo RT, Lopez S, Arias C. Structural polypeptides of simian rotavirus SA11 and the effect of trypsin. J Virol. 1981; 37:156–160. [PubMed: 6260970]
57. Estes MK, Graham DY, Mason BB. Proteolytic enhancement of rotavirus infectivity: molecular mechanisms. J Virol. 1981; 39:879–888. [PubMed: 6270356]

58. Ciarlet M, Estes MK. Human and most animal rotavirus strains do not require the presence of sialic acid on the cell surface for efficient infectivity. The Journal of general virology. 1999; 80(Pt 4): 943–948. [PubMed: 10211964]

59. Fuentes-Panana EM, Lopez S, Gorziglia M, Arias CF. Mapping the hemagglutination domain of rotaviruses. J Virol. 1995; 69:2629–2632. [PubMed: 7884915]

60. Isa P, Lopez S, Segovia L, Arias CF. Functional and structural analysis of the sialic acid-binding domain of rotaviruses. J Virol. 1997; 71:6749–6756. [PubMed: 9261399]

61. Hewish MJ, Takada Y, Coulson BS. Integrins alpha2beta1 and alpha4beta1 can mediate SA11 rotavirus attachment and entry into cells. J Virol. 2000; 74:228–236. [PubMed: 10590110]

62. Zarate S, Espinosa R, Romero P, Mendez E, Arias CF, Lopez S. The VP5 domain of VP4 can mediate attachment of rotaviruses to cells. J Virol. 2000; 74:593–599. [PubMed: 10623720]

63. Gilbert JM, Greenberg HB. Cleavage of rhesus rotavirus VP4 after arginine 247 is essential for rotavirus-like particle-induced fusion from without. Journal of virology. 1998; 72:5323–5327. [PubMed: 9573313]

64. Ciarlet M, Crawford SE, Estes MK. Differential infection of polarized epithelial cell lines by sialic acid-dependent and sialic acid-independent rotavirus strains. J Virol. 2001; 75:11834–11850. [PubMed: 11689665]

65. Haselhorst T, Fleming FE, Dyason JC, Hartnell RD, Yu X, Holloway G, Santegoets K, Kiefel MJ, Blanchard H, Coulson BS, von Itzstein M. Sialic acid dependence in rotavirus host cell invasion. Nat Chem Biol. 2009; 5:91–93. [PubMed: 19109595]

66. Svensson L. Group C rotavirus requires sialic acid for erythrocyte and cell receptor binding. J Virol. 1992; 66:5582–5585. [PubMed: 1380096]

67. Dormitzer PR, Sun ZY, Wagner G, Harrison SC. The rhesus rotavirus VP4 sialic acid binding domain has a galectin fold with a novel carbohydrate binding site. EMBO J. 2002; 21:885–897. [PubMed: 11867517]

68. Lobsanov YD, Gitt MA, Leffler H, Barondes S, Rini JM. Crystallization and preliminary X-ray diffraction analysis of the human dimeric S-Lac lectin (L-14-II). J Mol Biol. 1993; 233:553–555. [PubMed: 8411163]

69. Lobsanov YD, Gitt MA, Leffler H, Barondes SH, Rini JM. X-ray crystal structure of the human dimeric S-Lac lectin, L-14-II, in complex with lactose at 2.9-A resolution. J Biol Chem. 1993; 268:27034–27038. [PubMed: 8262940]

70. Yang RY, Rabinovich GA, Liu FT. Galectins: structure, function and therapeutic potential. Expert Rev Mol Med. 2008; 10:e17. [PubMed: 18549522]

71. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem. 2010; 31:455–461. [PubMed: 19499576]

72. Graham KL, Takada Y, Coulson BS. Rotavirus spike protein VP5* binds alpha2beta1 integrin on the cell surface and competes with virus for cell binding and infectivity. The Journal of general virology. 2006; 87:1275–1283. [PubMed: 16603530]

73. Basta HA, Sgro JY, Palmenberg AC. Modeling of the human rhinovirus C capsid suggests a novel topography with insights on receptor preference and immunogenicity. Virology. 2014; 448:176–184. [PubMed: 24314648]

74. Delos SE, Burdick MJ, White JM. A single glycosylation site within the receptor-binding domain of the avian sarcoma/leukosis virus glycoprotein is critical for receptor binding. Virology. 2002; 294:354–363. [PubMed: 12009877]

75. Shirato K, Miyoshi H, Goto A, Ako Y, Ueki T, Kariwa H, Takashima I. Viral envelope protein glycosylation is a molecular determinant of the neuroinvasiveness of the New York strain of West Nile virus. The Journal of general virology. 2004; 85:3637–3645. [PubMed: 15557236]

76. Botarelli P, Houlden BA, Haigwood NL, Servis C, Montagna D, Abrignani S. N-glycosylation of HIV-gp120 may constrain recognition by T lymphocytes. J Immunol. 1991; 147:3128–3132. [PubMed: 1717587]

77. Hause BM, Stine DL, Sheng Z, Wang Z, Chakravarty S, Simonson RR, Li F. Migration of the swine influenza virus delta-cluster hemagglutinin N-linked glycosylation site from N142 to N144

results in loss of antibody cross-reactivity. Clin Vaccine Immunol. 2012; 19:1457–1464. [PubMed: 22815146]

78. Lazdins I, Coulson BS, Kirkwood C, Dyall-Smith M, Masendycz PJ, Sonza S, Holmes IH. Rotavirus antigenicity is affected by the genetic context and glycosylation of VP7. Virology. 1995; 209:80–89. [PubMed: 7747487]

Research Highlights

- We built the RVC capsid structure model using a homology modeling based approach.
- We identified RVC VP7 surface exposed loops with potential antigenic properties.
- RVC VP4 model showed a reduced RVC VP4 foot in comparison to RVA VP4.
- We identified differences between N-linked glycosylation sites on RVA and RVC VP7.

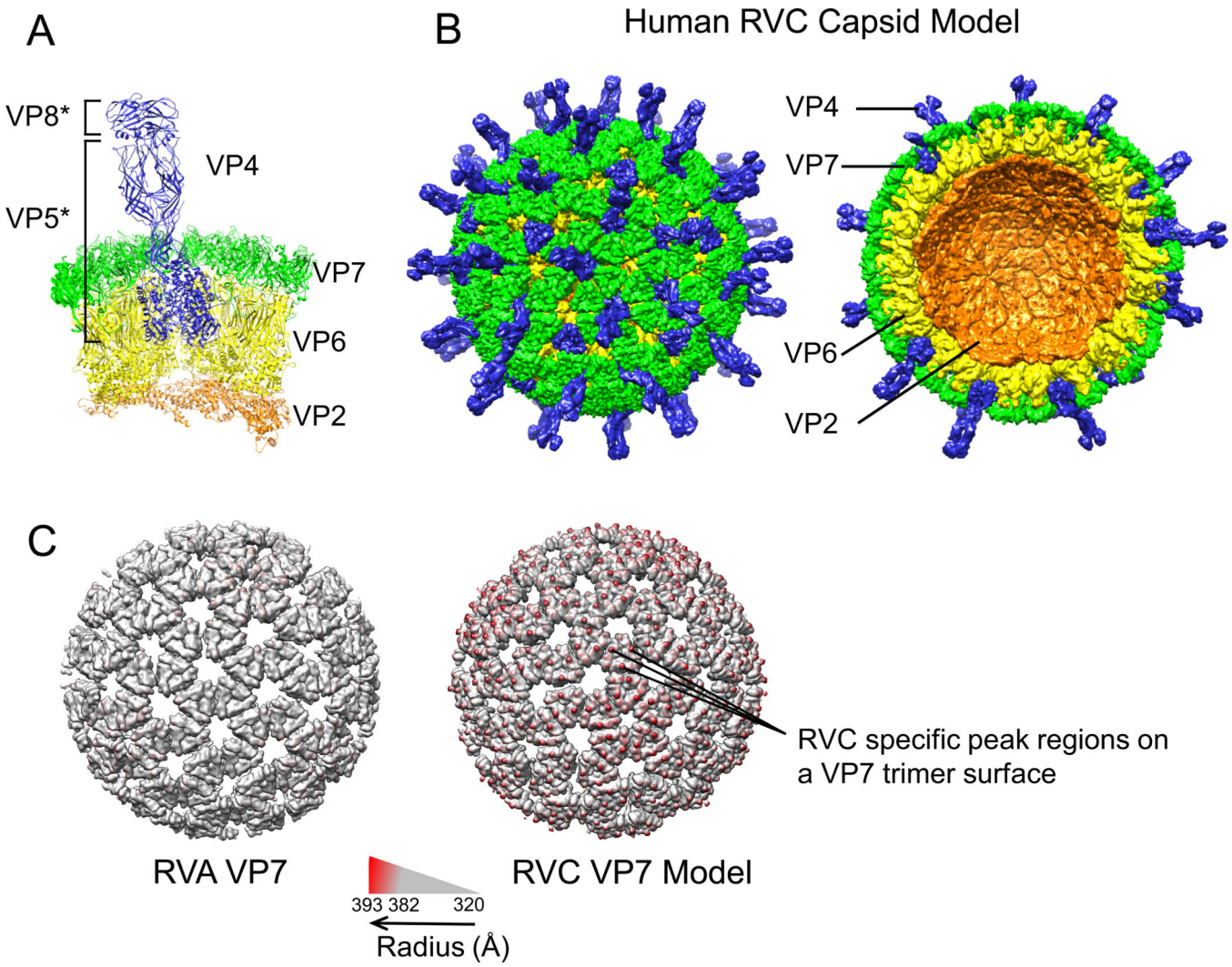

Figure 1.

Human RVC capsid model. (A) RVC 31-mer asymmetric unit model consisting of 3 VP4 (blue), 13 VP7 (green), 13 VP6 (yellow) and 2 VP2 (orange) proteins. VP4 head (VP8*) and body/stalk/foot (VP5*) illustrate products of trypsin cleavage. (B) Intact RVC icosahedral capsid model formed from the asymmetrical unit (left) and a cross-section of the capsid model (right). (C) RVA and RVC model VP7 layers colored according to radius from the center using Chimera. In order to emphasize the difference between the two VP7 layers, radii below 382 Å are shown in gray.

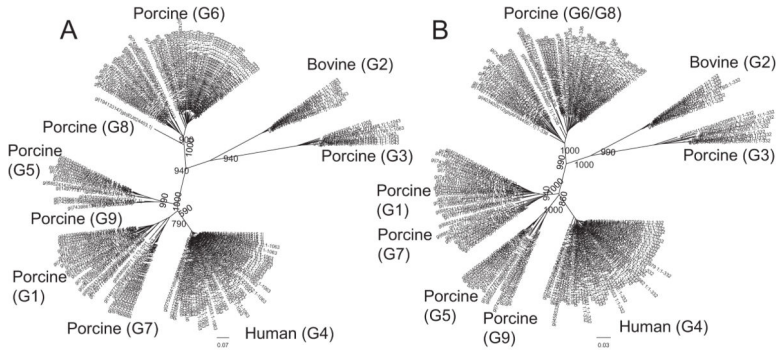

Figure 2.

Phylogenetic analysis of RVC VP7. (A) Phylogenetic tree constructed using 168 RVC VP7 nucleotide sequences. (B) Phylogenetic tree constructed using 168 RVC VP7 protein sequences. Corresponding G-types (Marthaler et al., 2013) are shown in parenthesis. Bootstrap values corresponding to the main nodes in both trees are shown.

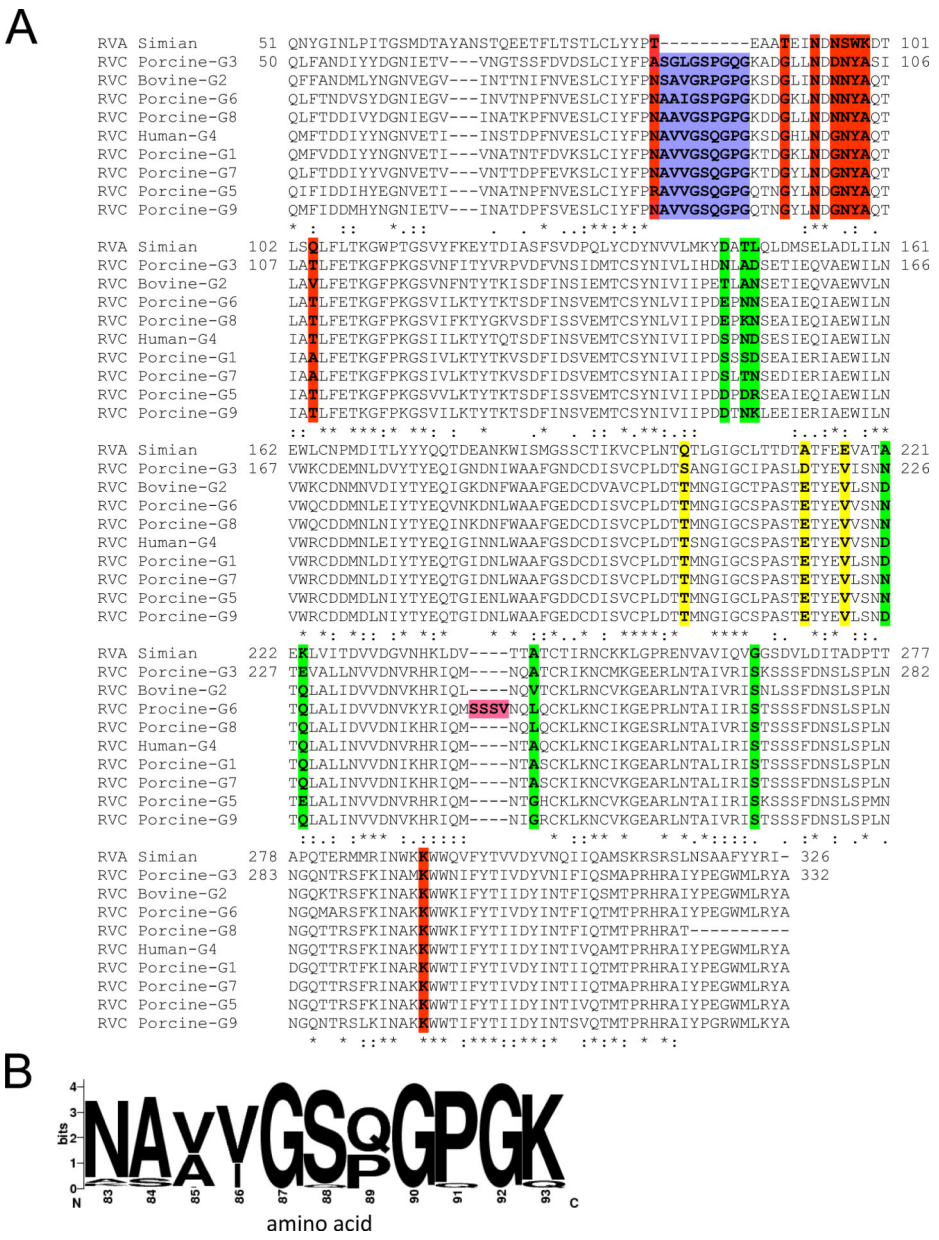

Figure 3.

Sequence comparisons of RVC VP7 proteins with RVA VP7. (A) Representative alignment of RVC VP7 proteins with simian RVA. Accession numbers for sequences are: ACC94320 for simian (RRV) RVA; YP_392513 (Bristol [G4]) for human RVC; AFR44611 (G1), AAC55009 (G3), BAQ15438 (G5), ABR28507 (G6), AJL35106 (G7), ACF33463 (G8) and BAQ15448 (G9) for porcine RVC; and Q65527 (Shintoku [G2]) for bovine RVC. RVA antigenic sites 7-1a (red), 71-b (yellow) and 7-2 (green), L9 insertion (blue) and L4a insertion (magenta) are indicated. (B) Sequence logo of RVC VP7 (amino acids 83–92) created using 166 sequences.

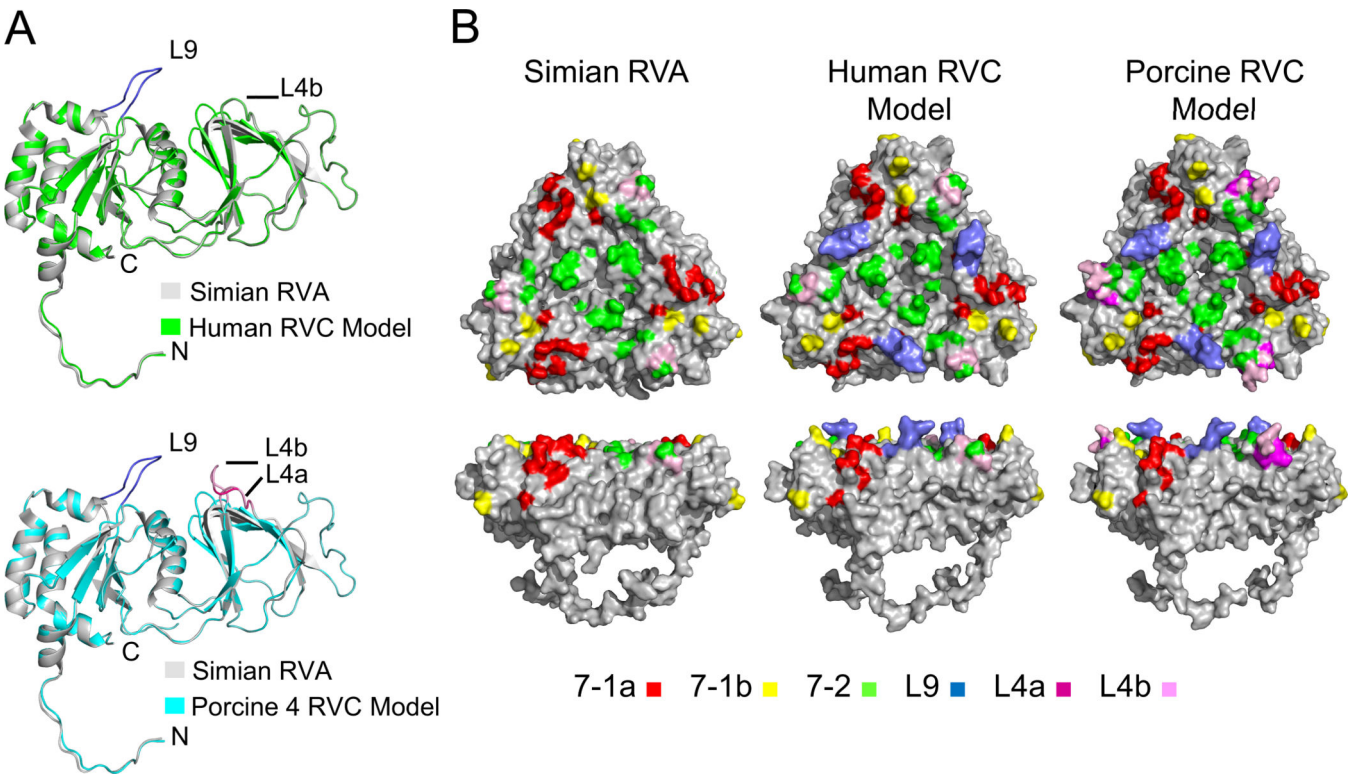

Figure 4.

Comparison of RVA VP7 and RVC VP7 model structures. (A) Alignment of human RVC (green, top) and porcine-G6 RVC (cyan, bottom) VP7 model structures to simian RVA (PDB ID 3IYU_A) (gray) VP7 structure. L9 and L4a insertions are colored in blue and magenta, respectively. L4b is shown in pink for porcine-G6 structure. (B) Surface representations of RVA and RVC VP7 model trimers. Antigenic regions selected by various mAbs mapped on simian RVA (left), and simian RVA antigenic regions projected on human (middle) and porcine-G6 RVC (right) models. Locations of L9, L4a, and L4b are also mapped on the surfaces. A color coding scheme is included at the bottom.

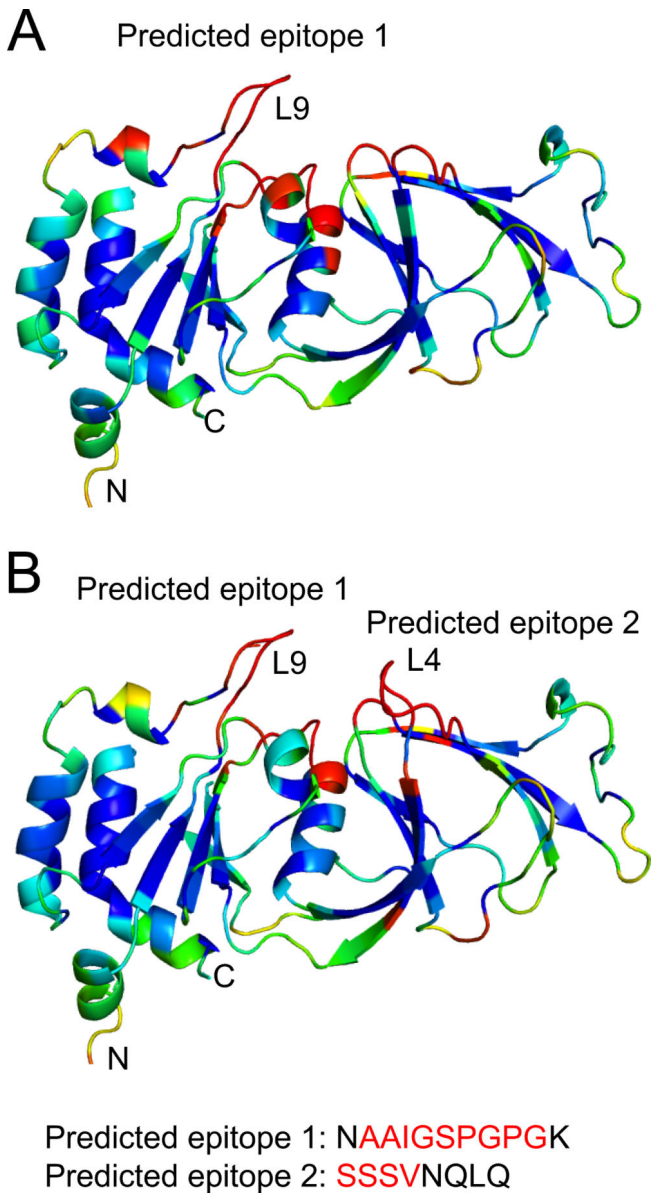

Figure 5.

Prediction of antigenic epitopes on RVC VP7 model surfaces by consensus scoring. Surface residues are colored from red to blue according to the predicted antigenicity (red, highest possibility; and blue, lowest possibility). (A) Human RVC VP7 model and (B) Porcine-G6 RVC VP7 model. L9 and L4a insertions are shown in red letters.

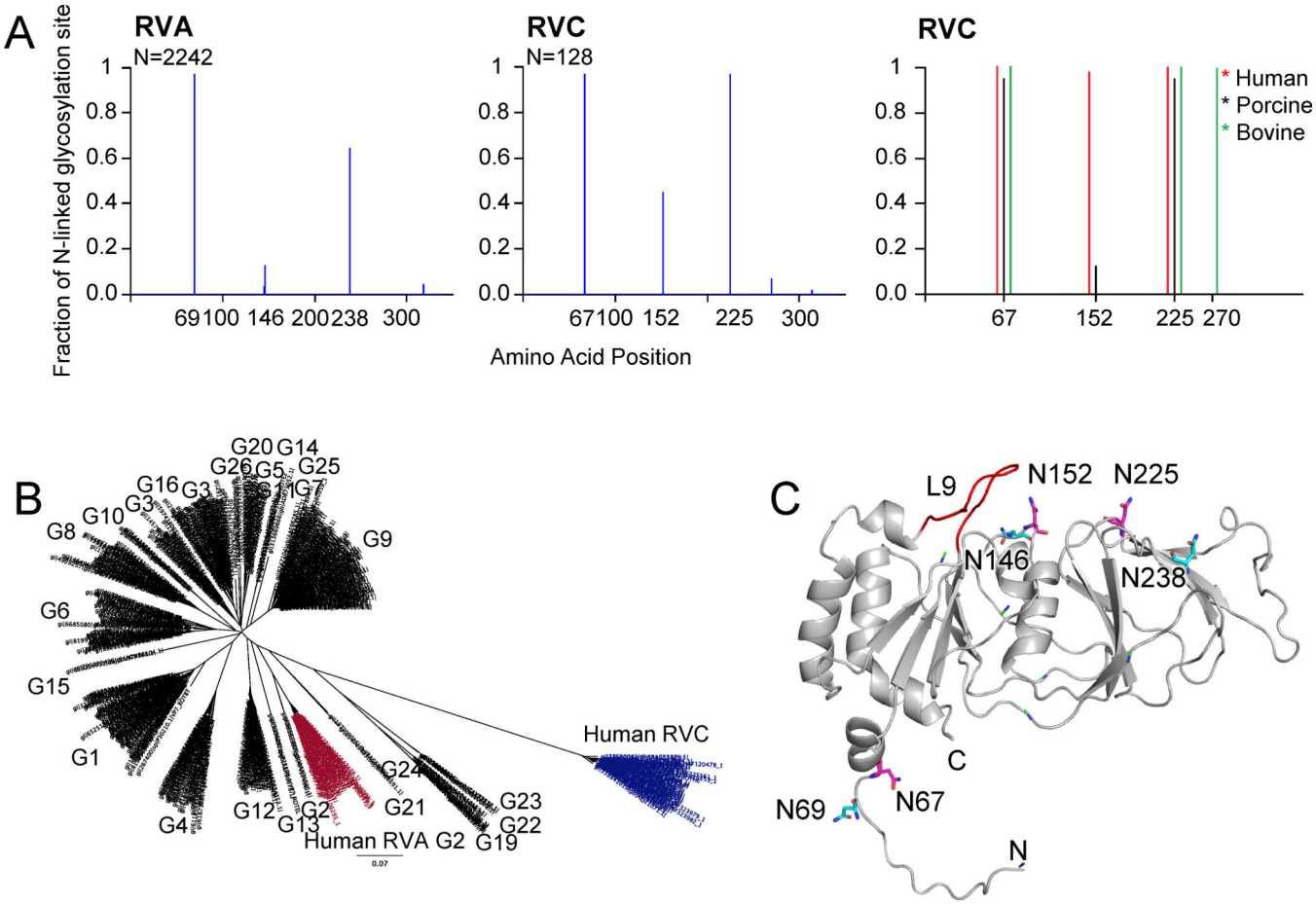

Figure 6.

Potential N-linked glycosylation sites of VP7 proteins. (A) Graphical representation of predicted glycosylation sites for RVA (left) and RVC (middle) VP7 proteins. Number of sequences used for the prediction (N) is shown on top left corner of the graphs. Distribution of most conserved N-linked glycosylation sites in human (red), porcine (blue) and bovine (green) RVC VP7 groups are also shown (left). (B) Phylogenetic tree of RVA and human RVC VP7 proteins. Human RVA G2-type VP7 sequences with 3 N-linked glycosylation sites are shown in red and human RVC VP7s are shown in blue. Corresponding G-types based on Matthijnssens et al., (2012) are also shown. (C) Mapping of potential conserved N-linked glycosylation sites on human RVC VP7 model. RVC conserved N-linked glycosylation sites are labeled and shown in magenta. RVA conserved N-linked glycosylation site locations are projected on the RVC model and are shown in cyan.

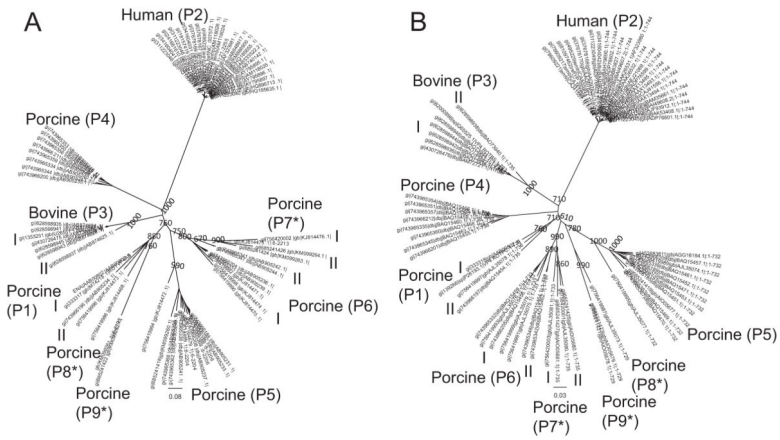

Figure 7.

Phylogenetic analysis of RVC VP4. (A) Phylogenetic tree constructed using 70 RVC VP4 nucleotide sequences. (B) Phylogenetic tree constructed using 70 RVC VP4 protein sequences. Corresponding P-types (Suzuki et al., 2015) are shown in parenthesis. P-types that differ from the previously published P-types are indicated with asterisks. Bootstrap values corresponding to the main nodes in both trees are shown.

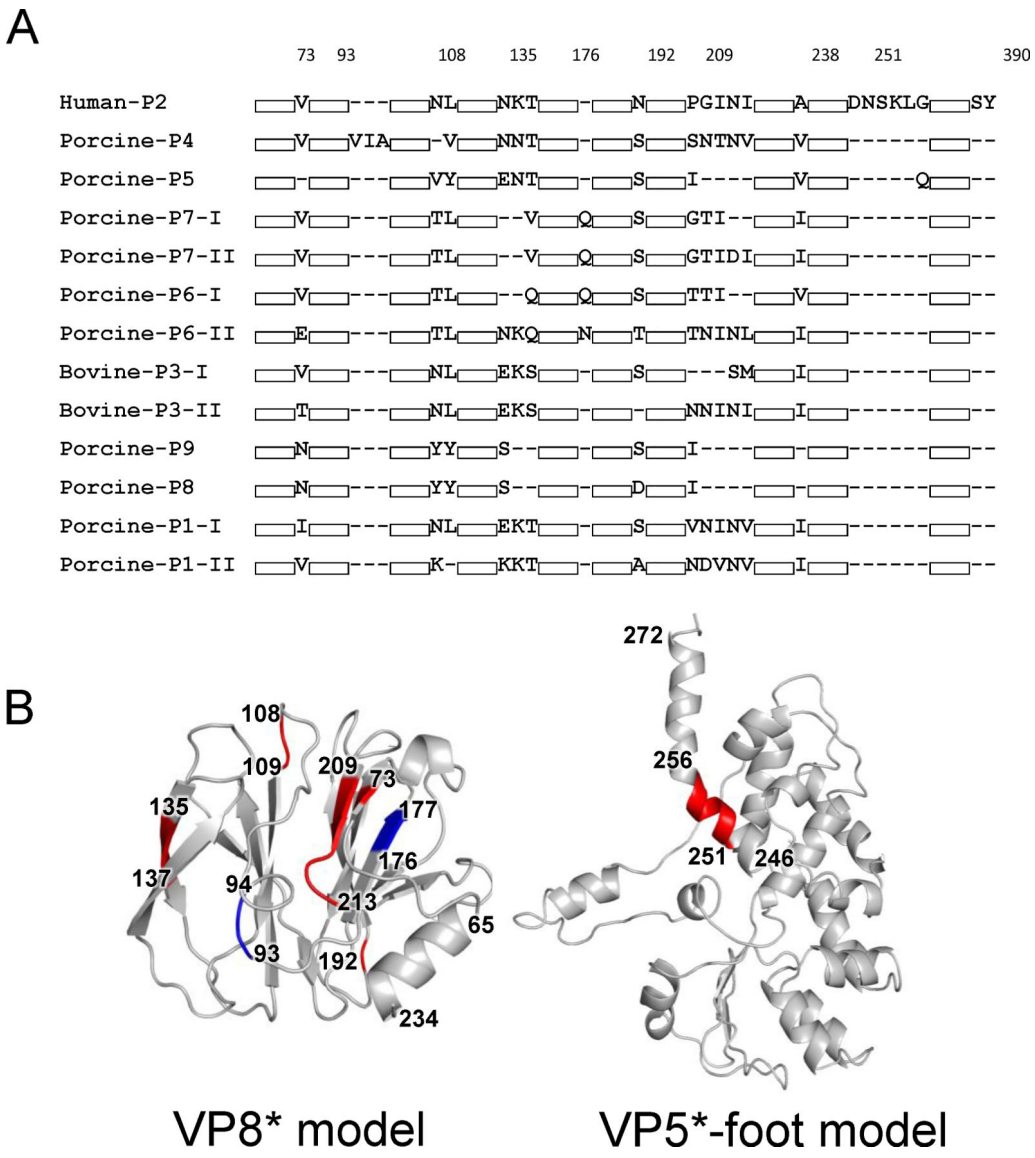

Figure 8.

Sequence and structure comparisons of RVC VP4 proteins. (A) Sequence alignment of representative VP4 proteins with different P-types showing sequence insertions and deletions (indels). For simplicity, regions without indels are shown as white boxes. Numbering of amino acids on top is done according to human RVC VP4 sequence. Accession numbers for sequences shown are: AAB00802, P1-1; BAQ15454, P1-2; YP_392514, P2; BAO73942, P3-1; BAO73940, P3-2; BAQ15460, P4; BAQ15452, P5; BAQ15456, P6-1; BAQ15464, P6-2; AJL35081, P7-1; AJL35080, P7-2; AJL35077, P8; AJL35073, P9; and AIO05679, P10. (B) Structural mapping of amino acid indels on human RVC VP4 with respect to other RVC VP4 proteins. Regions of insertions are shown in red and regions of deletions are shown in blue.

A

```
RVA Simian   1 MASLIYRQLLTNSYTVDLSDEIQEIGSTKTQNVTINPGPFAQTGYAPVNWGPGETNDSTTVEPVLDGPYQ 70
RVC Human    1 MASSLYAQLISQNYYSLGNEILSDQQTNKVVSDYVDAGNYTYAQLPPTTWGSGSILKSAFSTPEITGPHT 70

RVA Simian  71 PTTFNPP----VDYWMLLAPTAA----GVVVEGTNNTDRWLATILVEPNVTSETRSYTLFGTQEQITIAN 132
RVC Human   71 NTVIEWSNLINTNTWLLYQKPLNSVRLLKHGPDTYNSNLAAFELWYGKSGTTITSVYYNTINNQNKTHDA 140

RVA Simian 133 ASQTQWKFIDVVKTTQNGSYSQYGPLQSTPKLYAVMKHNGKIYTYNGKTPNVTTRYYSTTN------YDS 196
RVC Human  141 NSDCLILFWNEGSTQLEKQVVTFNWNVGGILIKPINSSRMRICMSGMENFNNDSFNWENWNHEFPRSNPG 210

RVA Simian 197 VNMTAFCDFYIIPREEESTCTEYINNGLPPIQNTRNIVPLALSARNIISHRAQANEDIVVSKTSLWKEMQ 266
RVC Human  211 ININMYTEYFLASS----DPYTYLKNLQQPTAKTVDMKMMKKMNDNSKLGDGPINVSNIISKDSLWQEVQ 276

RVA Simian 267 YNRDITIRFKFASSIVKSGGLGYKWSEISFKPANYQYTYTRDGEEVTAHTTCSVNGMNDFNFNGGSLPTD 336
RVC Human  277 YVRDITLQCKILSEIVKGGGWGYDYTSVTFKTVNHTYSYTRAGENVNAHVTISFNNVKERAY-GGSLPTD 345

RVA Simian 337 FVISRYEVIKENSYVYVDYWDDSQAFRNMVYVRSLAANLNSVI---CTGGDYSFALPVGQWPVMTGGAVS 403
RVC Human  346 FKIGRFDILDTDSYVYIDYWDDSEIFKNMVYVRDVRADIGGFQYSYSSEMSYYFQIPVGSYPGLHSSRLQ 415

RVA Simian 404 LHSAGVTLSTQFTDFVSLNSLRFRFRLTVEEPSFSITRTRVSRLYGLPAANPNNGKEYYEVAGRFSLISL 473
RVC Human  416 LVYDRCLLSQQFTDYAALNSLRFVFRVVSTSGWFITTGDINTRRVASGTGFAYSDGHVANTVGTISFISL 485

RVA Simian 474 VPSNDDYQTPITNSVTVRQDLERQLGELREEFNALSQEIAMSQLIDLALLPLDMFSMFSGIKSTIDAAKS 543
RVC Human  486 IPSNPNYQTPIASSSTVRMDLERKINDLRDDFNALASSVALSDILSLAMSPLTFSNLLESVPAITSSVKD 555

RVA Simian 544 MATSVMKKFKKSGLANSVSTLTDSLSDAASSISRGASIRSVGSSASAWTDVSTQITDVSSSVSSISTQTS 613
RVC Human  556 VAASVMKKFRSTKMFKKAAKQNYREFV----------------------IGDLLED------------ 589

RVA Simian 614 TISRRLRLKEMATQTEGMNFDDISAAVLKTKIDRSTQISPNTLPDIVTEASEKFIPNRAYRVINNDEVFE 683
RVC Human  590 -------VTNVARNNNSLNYSDITSAMMVSTTNRLQITDVDTFSEIVSRSADNFISNRSYRMIENNTVHE 652

RVA Simian 684 AGTDGRFFAYRVETFDEIPFDVQKFADLVTDSPVISAIIDFKTLKNLNDNYGISRQQAFNLLRSDPRVLR 753
RVC Human  653 IT-PTRRFSYDIKTLQQRNFDIDKFSKLASQSPVISAIVDFATIKAIRDTYGISDDIIYKLVASDAPTIL 721

RVA Simian 754 EFINQDNPIIRNRIEQLIMQCRL 776
RVC Human  722 SFINQNNPLIRNRITNLINQCKL 744
```

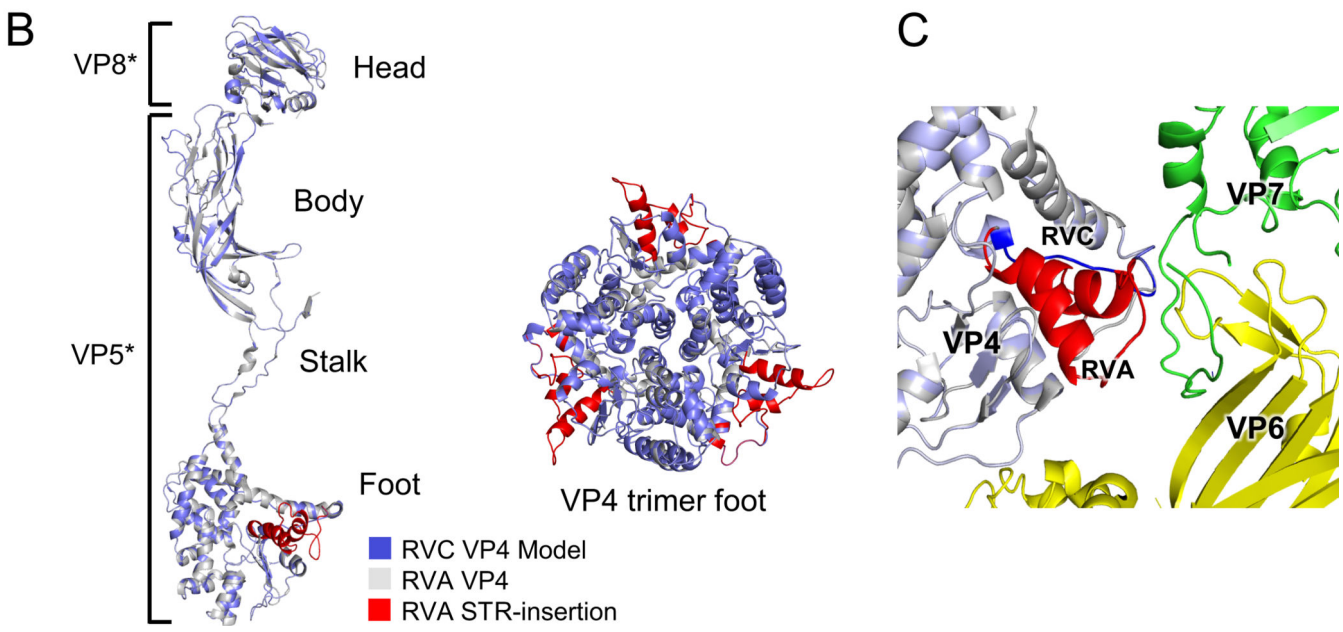

Figure 9.

Sequence and structure comparisons of RVA and RVC VP4 proteins. (A) Sequence alignment of human RVC VP4 (YP_392514) with simian RVA (ACC94315). Simian RVA sialic acid (SA) coordinating residues are shown in yellow. Trypsin cleavage sites are shown in green and integrin recognition peptide is shown in magenta. (B) Human RVC VP4 monomer structure model (blue) aligned with simian RVA VP4 structure (gray) is shown on left. VP8* (head) and VP5* (body, stalk and foot) regions are indicated. The 50 aa STR-insertion in the foot region of RVA VP4 is colored in red. Close-up view of the structural

alignment of simian RVA VP4 trimer foot (gray) with RVC VP4 foot (blue) as viewed from top (right). (C) Close-up view of the location of RVA STR-insertion in the capsid. VP7 is shown in green, VP6 is shown in yellow, RVA VP4 is shown in gray and STR-insertion is shown in red. RVC VP4 (blue) is superimposed on RVA VP4 and the shorter connecting region of RVC VP4 is shown in dark blue.

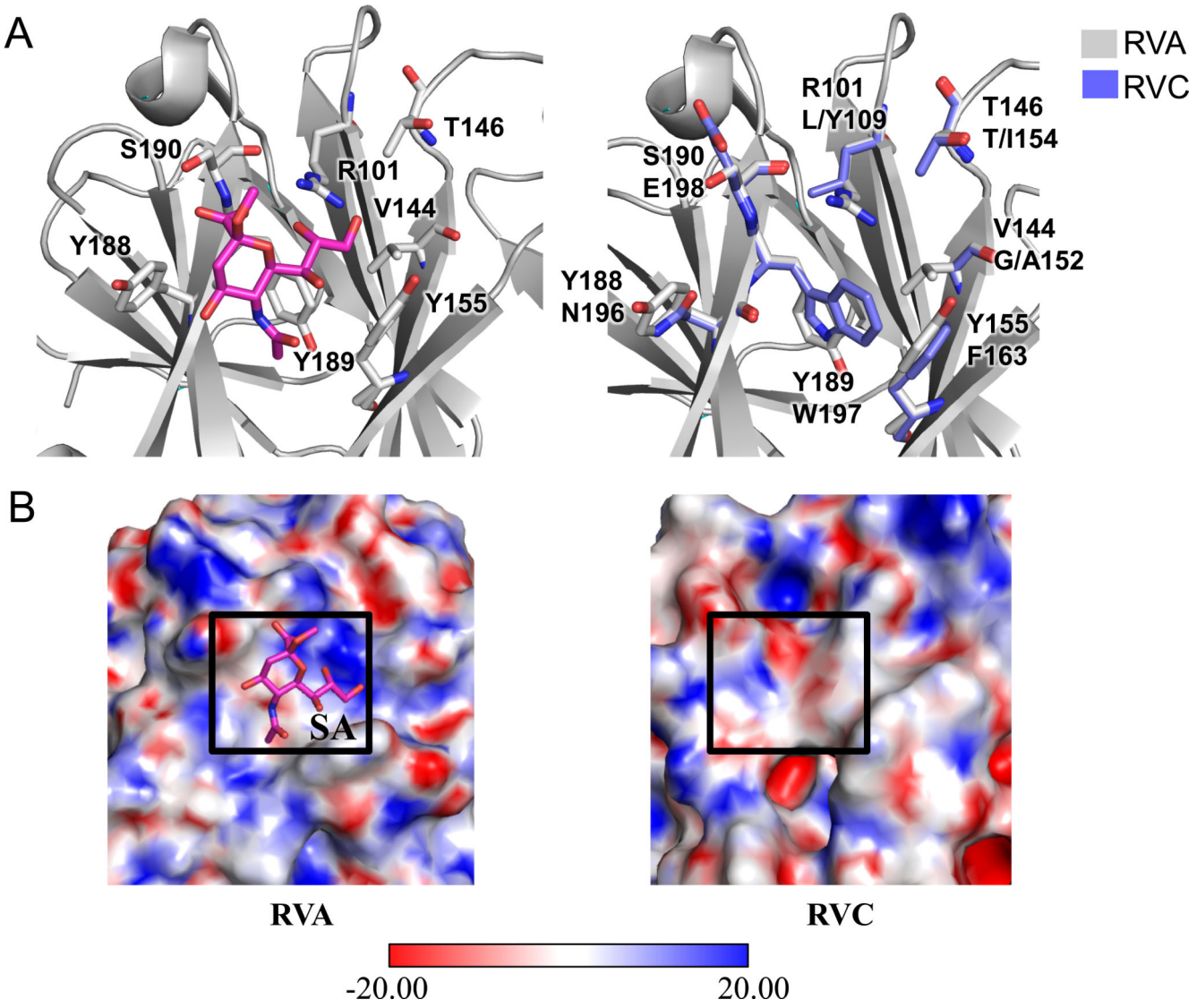

Figure 10.

Comparison of carbohydrate binding sites of RVA and RVC VP4 proteins. (A) Sialic acid (SA) binding site of simian RVA VP8* (PDB ID 1KQR) (left) and structural alignment of simian RVA VP4 SA binding site (gray) with human RVC VP4 model (blue, right). RVA residues coordinating SA are shown at the top and corresponding RVC residues are shown at the bottom. (B) Electrostatic surface potentials for simian RVA VP8* (left), and human RVC VP8* model (right). SA binding site is indicated with a box on both surfaces.

Table 1

Homology Model Evaluation Statistics

Model	% Identity	RMSD[1]	Q-Score	Ramachandran (%)	Verify3D
Human VP2	47.3	0.00	1.00	93.9[2]/3.8[3]/2.3[4]	0.36
Human VP4	31.9	0.54	0.84	94.4/3.3/2.3	0.27
Human VP6	43.0	0.16	0.98	99.4/0.3/0.3	0.40
Human VP7	35.5	0.11	0.96	95.9/2.9/1.3	0.37
Porcine-G6 VP7	36.0	0.11	0.94	99.5/3.3/1.7	0.37

[1] Values were calculated for Cα backbone

[2] Allowed regions

[3] Generously allowed regions

[4] Disallowed regions

Table 2

Percent Identity Matrix of RVC VP7 Proteins[1,2]

	G1	G2	G3	G4	G5	G6	G7	G8	G9
G1	91–100								
G2	74–75	98–100							
G3	69–73	74–75	90–100						
G4	85–87	73–74	71–72	96–100					
G5	83–85	72–74	69–72	84–85	88–100				
G6	78–79	78	71–72	80	77–79	92–100			
G7	88–90	74	71–73	83–87	82–85	80–81	91–100		
G8	77–78	76–78	70–71	79	75–79	87–89	79–80	100	
G9	84–88	75–76	69–73	85–86	86–89	79–80	84–87	80	92–100

[1] Percent identity matrix of RVC VP7 proteins forming different G-types were calculated using a Kimura-corrected pairwise alignment with a gap opening penalty of 6 bits and a gap extension penalty of 1 bit.

[2] Number of independent sequences used for each G-type is as follows: G1: 20; G2: 8; G3: 9; G4: 52; G5: 10; G6: 53; G7: 10; G8: 1 and G9: 5

Table 3

Percent Identity Matrix of RVC VP4 Proteins[1,2]

	P1	P2	P3	P4	P5	P6	P7	P8	P9
P1	85–100								
P2	73–74	96–100							
P3	77–81	72–74	90–100						
P4	78	71–72	74–76	91–100					
P5	75–76	69–70	74–76	75–76	84–100				
P6	80–81	71–72	78–79	76–77	74–82	90–100			
P7	76–78	70–71	75–77	73–76	72–77	82–85	83–100		
P8	75–76	71	76–77	77	76–78	78–79	75–77	100	
P9	75–76	68	73–75	73–74	73–74	74–76	76–79	82	100

[1] Percent identity matrix of RVC VP4 proteins forming different P-types were calculated using a Kimura-corrected pairwise alignment with a gap opening penalty of 6 bits and a gap extension penalty of 1 bit.

[2] Number of independent sequences used for each P-type is as follows: P1: 4; P2: 25; P3: 7; P4: 8; P5: 12; P6: 7; P7: 4; P8: 1, P9: 2.