







Novel fold of rotavirus glycan-binding domain predicted by AlphaFold2 and determined by X-ray crystallography

Liya Hu ¹, Wilhelm Salmen¹, Banumathi Sankaran ², Yi Lasanajak³, David F. Smith³, Sue E. Crawford⁴, Mary K. Estes ^{4,5} & B. V. Venkataram Prasad ^{1,4}

The VP8* domain of spike protein VP4 in group A and C rotaviruses, which cause epidemic gastroenteritis in children, exhibits a conserved galectin-like fold for recognizing glycans during cell entry. In group B rotavirus, which causes significant diarrheal outbreaks in adults, the VP8* domain (VP8*B) surprisingly lacks sequence similarity with VP8* of group A or group C rotavirus. Here, by using the recently developed AlphaFold2 for ab initio structure prediction and validating the predicted model by determining a 1.3-Å crystal structure, we show that VP8*B exhibits a novel fold distinct from the galectin fold. This fold with a β -sheet clasping an α -helix represents a new fold for glycan recognition based on glycan array screening, which shows that VP8*B recognizes glycans containing N-acetyllactosamine moiety. Although uncommon, our study illustrates how evolution can incorporate structurally distinct folds with similar functionality in a homologous protein within the same virus genus.

¹Verna and Marrs McLean Department of Biochemistry and Molecular Biology, Baylor College of Medicine, Houston, TX, USA. ²Berkeley Center for Structural Biology, Molecular Biophysics and Integrated Bioimaging, Lawrence Berkeley Laboratory, Berkeley, CA, USA. ³Emory Glycomics and Molecular Interactions Core (EGMIC), Emory University School of Medicine, Atlanta, GA, USA. ⁴Department of Molecular Virology and Microbiology, Baylor College of Medicine, Houston, TX, USA. ⁵Department of Medicine, Baylor College of Medicine, Houston, TX, USA. ✉email: lhu@bcm.edu; vprasad@bcm.edu

Rotaviruses are non-enveloped icosahedral double-stranded RNA (dsRNA) viruses belonging to the *Reoviridae* family¹. These viruses exhibit enormous genetic and serological diversity. Based on the sequence and antigenic differences of the capsid protein VP6, they are classified into ten different species or groups (A–J)^{2,3}. Group A, B, C, and H rotaviruses infect both humans and animals^{4,5}. Epidemiologically, groups A, B, and C are the best characterized. While group A rotaviruses (RVA) and to a lesser extent group C rotaviruses (RVC) are the causative agents of most gastroenteric infections worldwide, the group B rotaviruses (RVB) have been associated with large epidemic outbreaks of severe gastroenteritis in China^{6,7} and sporadic infection in several countries^{8–10}. Unlike RVA, which infects mainly young children, RVB causes cholera-like severe diarrhea predominantly in adults, although children can be infected as well⁸. Antibodies to the RVB have been detected in people in developed countries such as the USA, Canada, and the UK^{11,12}, indicating a broader prevalence of RVB. RVB also infects animals and caused recent hemorrhagic diarrheal outbreaks in piglets and foals, resulting in significant economic impact and posing a threat of potential zoonotic transmission^{13,14}. Considering that these rotaviruses have segmented dsRNA genomes, with a propensity for evolving by gene reassortment from co-infections and mutations, the potential for the emergence of new variants that can cause severe epidemics cannot be discounted. A case in point is the current ongoing global COVID-19 pandemic caused by SARS-CoV-2¹⁵.

Like RVA, which remains the best characterized thus far, and RVC, the genome of RVB consists of 11 dsRNA segments that encode 11 proteins¹⁶. Cryo-EM reconstructions of RVA have shown that trimers of VP4 form 60 protruding spikes attached to the outer VP7 and middle VP6 layers of the triple-layered capsid^{17–19}. The proteolytic treatment of VP4, which significantly enhances infectivity, results in two fragments, VP8* and VP5*, that remain associated with the virion. Sequence comparison of the structural proteins encoded by different rotavirus groups shows that the VP8* domain of the spike protein VP4 is the most variable²⁰. Extensive structural studies have shown that the galectin-like VP8* of human RVA and RVC (VP8*A and VP8*C) recognize various cellular glycans in a genotype-dependent manner (Fig. 1a, b)^{21–29}. The VP8* of human RVA exhibits genotype-dependent glycan specificity by recognizing different histo-blood group antigens (HBGA)³⁰. Differential recognition of HBGA has provided a possible rationale for why some RVAs specifically infect neonates, and some cause sporadic outbreaks while others infect a wider population^{22–24,31}. Unlike the VP8*A and VP8*C, the structure of VP8*B and its glycan specificity have not been characterized. The VP8*B shares no sequence identity with either VP8*A or VP8*C (Supplementary Fig. 1a and Supplementary Table 1), which could differentially impact not only the structure but also the glycan-binding properties. Here we show by determining the crystal structure of VP8*B, using a model predicted using the recently developed AlphaFold2, that the VP8*B exhibits a fold with a twisted β -sheet clasping an α -helix that is entirely different from VP8*A or VP8*C. Our glycan array screening and in silico docking analysis show VP8*B recognizes glycans containing N-acetylglucosamine exemplifying how viruses can evolve by incorporating structurally distinct modules with similar functionality.

Results

The AlphaFold2 model of VP8*B reveals a novel fold. To understand the structure and the glycan specificity of VP8*B, we undertook crystallographic studies of VP8*B and performed glycan array screening. As a representative VP8*B, the amino

acid sequence of a human RVB, isolated from a gastroenteritis outbreak in India³², was expressed and purified for structure determination by X-ray crystallography. The VP8*B crystals diffracted to 1.3-Å resolution. Consistent with the low sequence identity with either VP8*A or VP8*C (Supplementary Fig. 1a and Supplementary Table 1), our attempts to find a molecular replacement (MR) solution using their structures or the structural models predicted from computational programs, such as trRosetta³³ and I-TASSER³⁴, was unsuccessful. Instead of using the traditional single anomalous dispersion (SAD) phasing method with crystallization of the selenium-methionine substituted recombinant VP8*B, we used the recently developed AlphaFold2 to generate a suitable search model^{35,36}.

The predicted AlphaFold2 models were significantly different from any of the previously determined experimental structures of RV VP8*, potentially representing a divergent novel fold for a glycan-binding protein (Fig. 1c). In the AlphaFold2 model, the residues 83–192 fold into several β -strands forming a twisted β -sheet clasping a central α -helix, while the N- and C-terminal residues (aa 65–82 and aa 193–233) projecting away from this fold are flexible. To examine if a similar novel fold is also predicted for the VP8*B of other RVBs, we used the primary sequences of murine, bovine, and porcine RVB (Fig. S1a). The predicted fold for these sequences, despite only ~27–33% sequence identity, is the same as that predicted for VP8*B of human RVB (Supplementary Fig. 1b and Supplementary Table 1), suggesting that group B VP8* has diverged significantly from the well-characterized galectin-like fold of the VP8*A and VP8*C.

Crystal structure of VP8*B. We used the well-ordered region of VP8*B (residues 83–192) in the AlphaFold2-predicted fold as a search model for MR, and determined the crystal structure of native VP8*B at 1.3-Å resolution (Table 1). The crystallographic asymmetric unit contains two VP8*B molecules in addition to a short peptide (Fig. 2a). These two molecules in the asymmetric unit of the crystal structure superimposed with a Ca RMSD of 0.037 Å using Secondary Structure Matching (SSM) superpose in COOT³⁷. Although we purified and crystallized the recombinant VP8*B containing residues 65–233 of VP4, only the residues F78–N202 are observed in the crystal structure. The electron density of the N- and C-terminal residues are not observed due to their flexibility or a proteolytic cleavage that may have occurred during crystallization. The small peptide with four residues could be the protease cleavage product co-crystallized with the globular portion of the VP8*B. The electron density of the side chains does not match any residues of the N or C terminus of VP8*B, possibly due to heterogeneity of the short peptide and the lack of sequence specificity of the peptide-binding (Supplementary Fig. 2). The VP8*B structure contains seven β -strands and one α -helix (Fig. 2b, c). Five of these β -strands form a twisted antiparallel β -sheet that surrounds a central α -helix. In the primary sequence, this central α -helix is between the residues that form β 4 and β 5 strands.

To examine if this fold is unique, we used DALI server³⁸ to compare with other structures in the Protein Data Bank (PDB). In this comparison, no structures showed strong similarities. The two structures that showed a marginal similarity, with Z-scores of 3.6 and 3.5, respectively, are exo-inulinase and bacteriophage T4 gene product 9 (gp9) (Fig. 2d, e). Although these structures showed a similar disposition of the antiparallel β -sheet, they both lacked the central α -helix in the VP8*B structure. The RMSD between the matching 66 Ca atom pairs in the β -sheet region of the exo-inulinase is 2.7 Å, whereas the RMSD between 76 Ca atom pairs with gp9 is 3.6 Å (Fig. 2d, e).

Surprisingly, this DALI search did not identify similarity with either VP8*A or VP8*C structures deposited in the PDB,

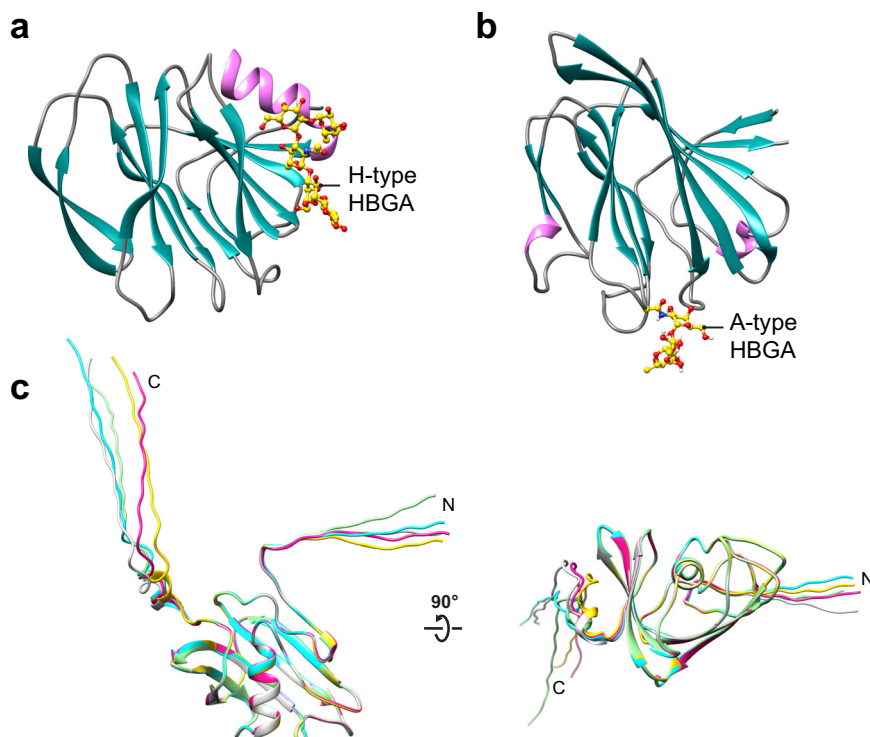


Fig. 1 **Ab initio modeling a human group B rotavirus VP8* with AlphaFold2 reveals a novel fold.** **a, b** Representative VP8* crystal structures: VP8*A in complex with H-type HBGA pentasaccharide (PDB ID: 5VX5) and VP8*C in complex with A-type HBGA trisaccharide (PDB ID: 5ZHO). The β -sheets and α -helices are colored in sea green and pink, respectively. **c** Top 5 ranked predicted models of VP8*B shown as ribbon representations in different colors.

Table 1 Data collection and refinement statistics.

RVB VP8* (PDB ID: 7RSW)

Data collection	
Space group	P 1 2 ₁ 1
Cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	31.277, 116.257, 31.312
α , β , γ (°)	90, 103.54, 90
Resolution (Å)	50.0–1.32 (1.34–1.32)
<i>R</i> _{merge}	0.04 (0.262)
CC1/2	0.995 (0.844)
<i>I</i> / σ <i>I</i>	19.03 (2.31)
Wilson B-factor	13.24
Completeness (%)	89.4 (61.4)
Redundancy	1.9 (1.4)
Refinement	
Resolution (Å)	30.41–1.32 (1.36–1.32)
No. reflections	41,757 (16,75)
<i>R</i> _{work} / <i>R</i> _{free} (%)	15.74 (22.41)/20.84 (34.18)
Number of atoms	
Protein	1,914
Water	285
B-factors	
Protein	16.99
Water	34.01
R.M.S. deviations	
Bond lengths (Å)	0.005
Bond angles (°)	0.84
Ramachandran favored (%)	98.79
Ramachandran allowed (%)	1.21
Ramachandran outlier (%)	0

Statistics for the highest-resolution shell are shown in parentheses.

although the galectin-like fold in these structures exhibits twisted antiparallel β -sheet domains. We then examined if the β -sheet in the VP8*B could be aligned with either of the two β -sheets in the galectin-like fold of VP8*A or VP8*C structures. We found that the β -sheet structure of VP8*B is indeed unique, and shared no similarity with the β -sheets in VP8*A or VP8*C

structures indicating the evolutionary path of VP8*B fold is distinct.

Comparison of the experimental and AlphaFold2 models of VP8*B. Having determined the structure of the VP8*B, we examined how closely the experimentally determined structure of VP8*B compared with the AlphaFold2-predicted model. Structural comparison of the two shows a high degree of similarity between the models with a C α RMSD of 0.398 Å and 0.403 Å with the two VP8*B molecules in the asymmetric unit, respectively (Fig. 3a, b). Further inspection showed that the sidechain orientations also match well except for His131 (Fig. 3c, d). The sidechain orientation of this residue is likely influenced by the intermolecular contacts and solvent in the crystal. In the experimental structure, the side chain of His131 hydrogen bonds with Gln195 within the same molecule and Glu120 of the neighboring symmetry-related molecule (Fig. 3d). There is also a water molecule that hydrogen bonds with His131 sidechain and the main chain carbonyl oxygen atom of Cys198. In the AlphaFold2 model, the side chain of His131 residue orients differently, and the same orientation would clash with the water molecule observed in the crystal structure.

AlphaFold model of full-length VP4B. The accuracy of the AlphaFold2-predicted fold of VP8*B prompted us to investigate other regions in the VP4 spike of RVB, particularly in comparison with the VP4 structure of RVA (Supplementary Fig. 3a). The structural organization of the VP4 spike in the RVA virion is well characterized by cryo-EM studies^{17,18}. In addition to the VP8* domain, the remainder of the spike protein VP4 in RVA consisting of VP5* is described as having a central body and a foot region that is buried inside the VP7 and VP6 layers at one of the channels of the icosahedral capsid (Supplementary Fig. 3b). The residues (248–479) that form the central body of the spike fold into a β -

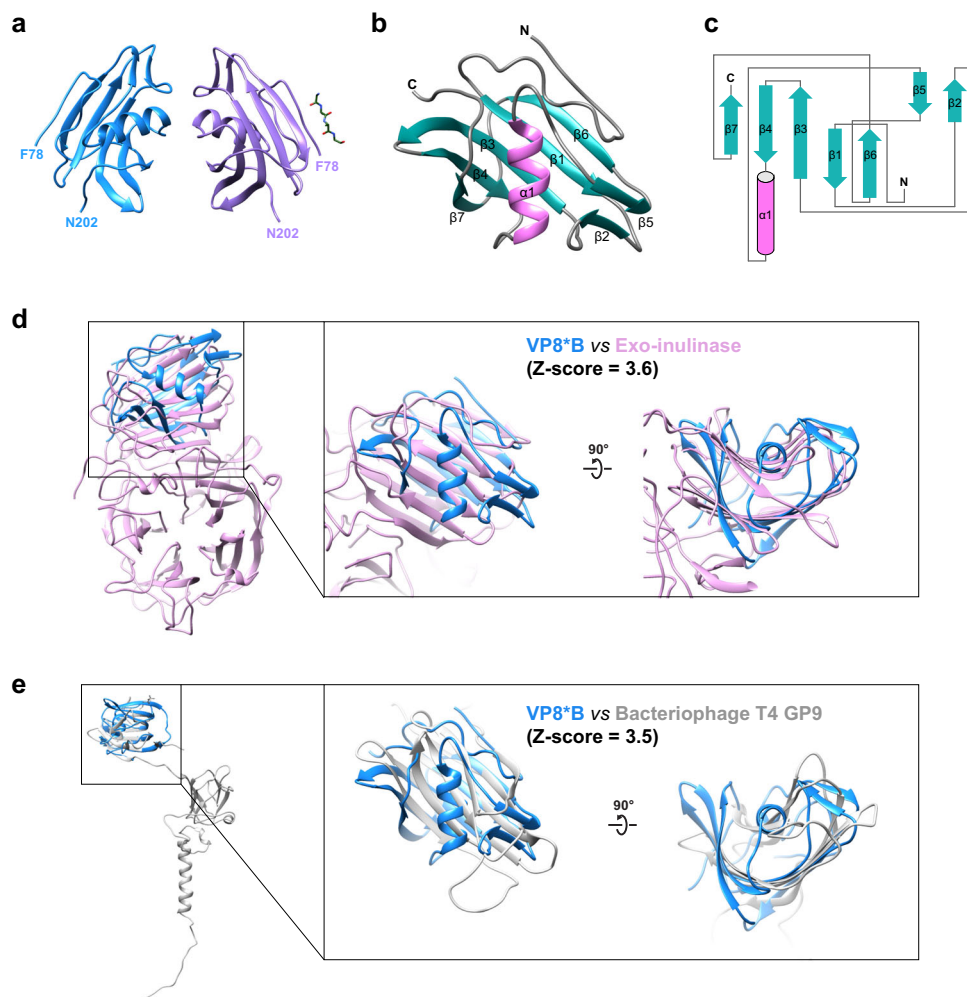


Fig. 2 Crystal structure VP8*B. **a** Crystal structure of VP8*B with chain A and B colored blue and purple, respectively. The co-crystallized peptide is shown as a stick model. **b** Monomer of VP8*B with β -strands and α -helix colored in sea green and pink, respectively, showing the β -sheet wrapped around the helix. **c** The 2D topology diagram of VP8*B. The β -strands of the helix are labeled in **(b)** and **(c)**. **d, e** Structural comparison of VP8*B with exo-inulinase and bacteriophage T4 gene product 9. The Dali Z-scores are shown in the figures.

barrel domain, whereas the residues that constitute the foot region, composed of residues (491–776), have α -helices as well as β -strands. The trimeric organization of the VP4 spike in RVA is unique in that the central body of the VP4 spike is formed by dimeric interaction between the β -barrel regions of the two subunits. The β -barrel domain of the other subunit, in which the VP8* domain is flexible and not observed in the spike structure, lies between the central dimeric part and proximal foot region. All three subunits of the VP5* contribute to the foot region and interact with three N-terminal α helices of VP8*. The flexible segments of the N-terminus of VP4A run between the two β -barrel domains and connect the N-terminal α helices and the lectin domain.

We used the full-length sequence of the VP4B to predict the structure of VP5*B using AlphaFold2 (Supplementary Fig. 3c). In this prediction, the VP8*B has the same novel fold as predicted by using the VP8*B sequence alone. Evolutionary conservation analysis using the ConSurf server³⁹ shows that VP8*B is the most variable domain within VP4 (Supplementary Fig. 3d). The residues (214–465) of the VP5*B fold into a β -barrel structure with the same directionality and the disposition of the β -strands as observed in the VP5*A except for the loop regions. The predicted β -barrel of VP5*B matches that of VP5*A with an RMSD of 3.4 Å (Supplementary Fig. 3f). However, the foot

domain of VP5*B, despite having a similar distribution of α -helices and β -strands, superimposed with the corresponding region in VP5*A with a high RMSD of 20.1 Å (Supplementary Fig. 3g), indicating the potential limitations of AlphaFold2 in the ab initio modeling of oligomeric multidomain proteins. Recent cryo-EM studies showed that VP4A can rearrange from an ‘upright’ to a ‘reversed’ conformation where the VP5* foot domain is embedded in the membrane¹⁷, suggesting that the structure prediction of the VP5* foot domain in the ‘reversed’ conformation should also consider the physical and chemical characteristics of the membrane environment.

VP8*B binds to glycans containing an N-acetylglucosamine. To investigate if VP8*B, despite its novel fold, binds to glycans and exhibits any specificity as its counterpart in RVA and RVC, both of which bind to HBGA, we performed high-throughput glycan array screening with the recombinant GST-tagged VP8*B protein^{21–24,29}. The glycan array screening shows that VP8*B specifically recognizes glycans containing an N-acetylglucosamine (LacNAc) motif (Fig. 4a, b and Supplementary Table 2), a precursor disaccharide for HBGA synthesis and a universal component of N- and O-glycans and glycolipids⁴⁰. The VP8* of human-bovine reassortant neonate-specific human P[11] RVA

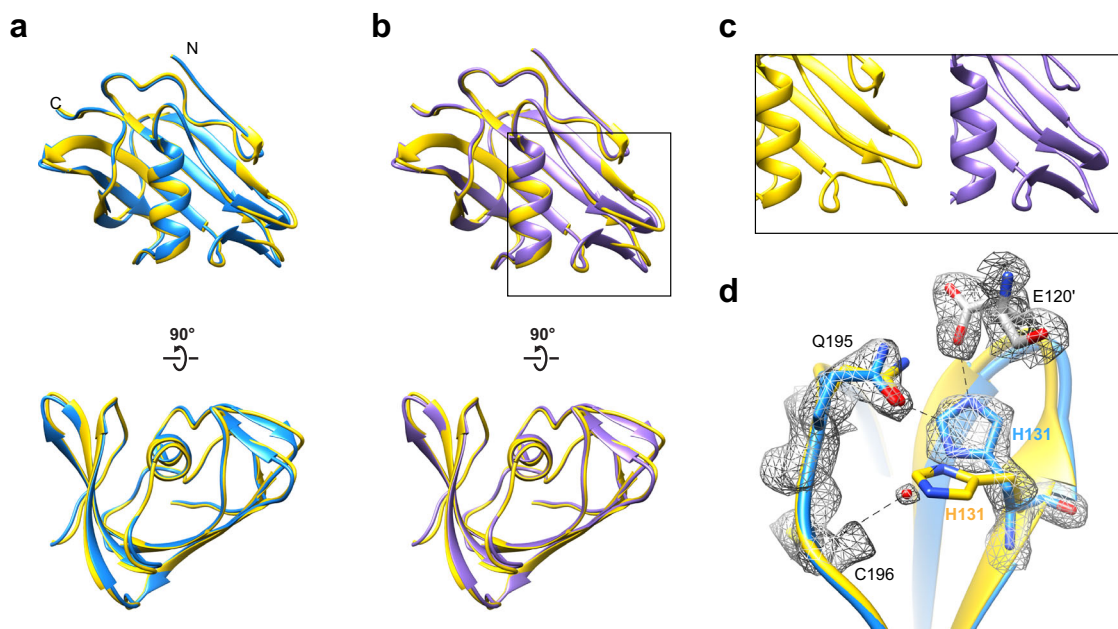


Fig. 3 Structural comparison of VP8*B crystal structure with AlphaFold2 prediction. **a, b** Superimposition of chain A and B of VP8*B in the crystallographic asymmetric unit with the AlphaFold2-predicted model. Crystal structure of RVP8*B with chains A and B colored in blue and purple, respectively. The residues 72–202 of the highest-ranked AlphaFold2 model is shown as a yellow ribbon diagram. The N- and C-termini of VP8*B are labeled. **c** The close-up view of the region in the black box in **(b)**. The superimposed models are shown side-by-side for clarity. **d** Example of structural difference between the AlphaFold2 model (yellow) and crystal structure (chain A, blue). The hydrogen bonds are shown as black dashed lines. The experimental 2Fo-Fc map of the residues is shown as gray mesh.

and the bovine P[11] RVA also recognize poly-LacNAc glycans^{22,31} for cell attachment, suggesting the possibility that LacNAc could also be a cell attachment factor for RVB.

To understand the molecular interactions between VP8*B and LacNAc disaccharide, we performed molecular docking using AutoDock Vina⁴¹. The docking pose with the lowest binding free energy of -5.7 kcal/mol shows that the common precursor motif of the selected glycans binds to a shallow pocket on VP8*B (Fig. 4c and Supplementary Fig. 4). Eight VP8*B residues are involved in a network of interactions with LacNAc. For example, K153, Y157, T160, S193, and Q195 form hydrogen bonds with LacNAc. Interestingly, the side chain of Q195 also interacts with His131 that is present in different conformations in the predicted and experimental models, suggesting that subtle changes in solution may affect glycan-binding during virus infection.

Discussion

Although it is common that under evolutionary pressures, a functionally homologous protein within the same genus in a virus family accrues mutations within a conserved polypeptide fold, it rarely evolves to adopt an entirely different fold. Such is the case with the VP8* domain of RVB, which adopts a new fold without any similarities to the galectin-like fold conserved in the VP8* of RVA and RVC. All these viruses infect various mammalian species, including humans causing severe diarrheal outbreaks. However, the difference is that RVA and RVC infect predominantly children under the age of 5, whereas RVB, also known as Adult Diarrhea Rotavirus (ADRV), infects predominantly adult populations. It is unclear whether this distinction alone explains such a unique fold of VP8*B. It likely represents a parallel evolution from a different ancestral animal origin from that of RVA and RVC. AlphaFold2 shows that this novel fold is conserved in other strains of RVB as well. Despite the novel fold, based on our glycan screening study, it is likely

that VP8*B, similar to the VP8* of RVA and RVC, retains the same functionality of mediating the initial cell attachment by recognizing specific glycans in the gut. Although LacNAc is one of the common glycans in the human and animal gut, the significance of VP8*B specificity to this glycan needs further infectivity-based studies, which are currently difficult as the RVB is not yet conducive to cell culture.

Another important aspect of our studies that should be underscored is the accuracy with which AlphaFold2 predicted the fold of VP8*B, which provided MR search model for our structure determination. In addition to the overall fold, AlphaFold2 also predicted the sidechain orientation with high accuracy when compared with the experimental structure. As exemplified in our studies, such accurate *ab initio* prediction will be an asset to structural biologists by providing partial or full MR search models for rapidly determining the structures of those proteins without suitable homologous structures. In addition to the VP8*B structure, AlphaFold2 predicted that VP5*B has a similar β -barrel fold as VP5*A, indicating possible conservation of this region. However, the foot region was not as well predicted by AlphaFold2, which as in the RVA may be influenced by the trimeric nature and potential interactions with the outer layer proteins in RVB. Further validation of the VP5* region of the VP4 spike in RVB requires a cryo-EM structure determination which has to await a successful cell culture adaptation of RVB.

Methods

ab initio modeling of RVB VP8* with AlphaFold2. The full AlphaFold v2.0 (AlphaFold2) pipeline was obtained from DeepMind and installed on a local workstation³⁶. The VP8*B amino acid sequence (residues 65–223) of human group B rotavirus strain NIV-094456 isolated from gastroenteritis outbreaks in India in 2009 was used as the template³². The VP8*B structure was predicted using the default setting with “ $-\text{max_template_date}=2020-05-14$ ” and “ $-\text{preset}=casp14$ ” which runs with all genetic databases and with eight ensemblings. The AlphaFold2 models of murine, bovine, and porcine VP8*B and human VP4B were predicted using the same protocol.

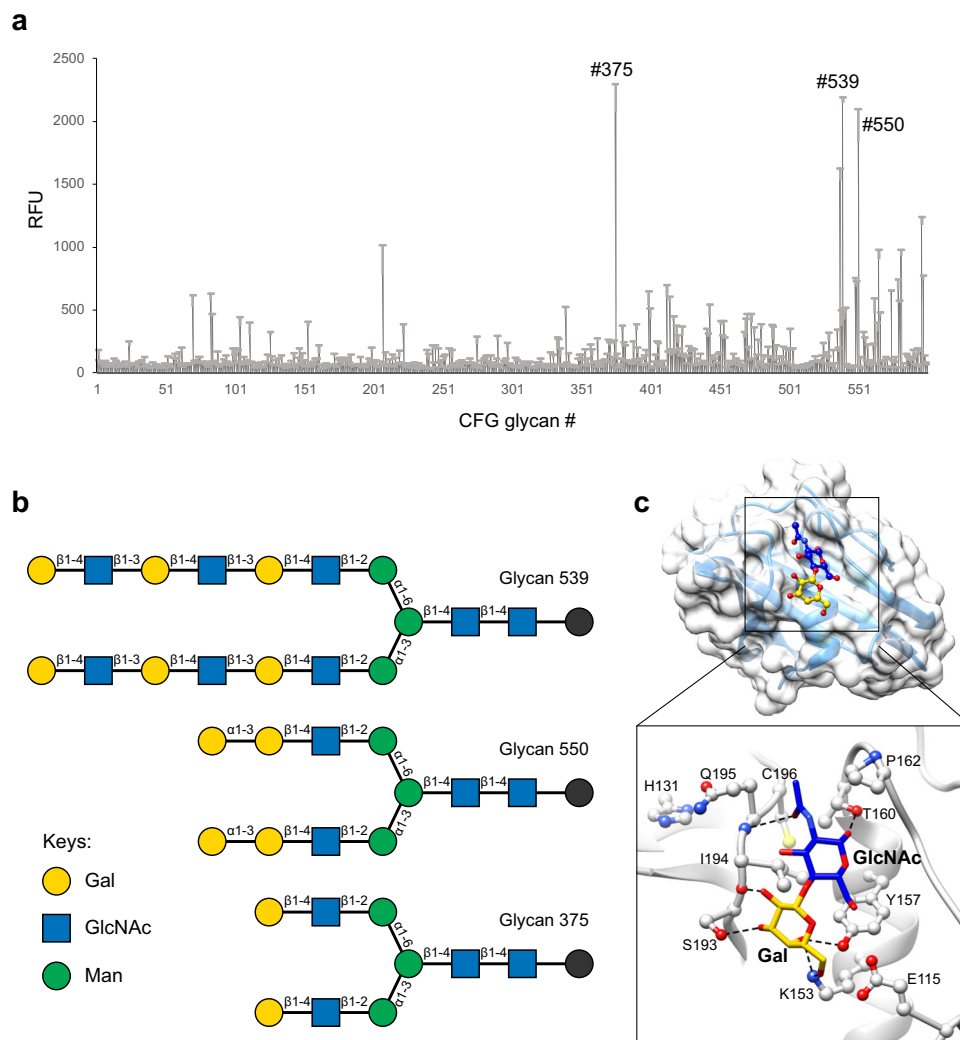


Fig. 4 Glycan array of VP8*B and molecular docking of the selected glycan. **a** The glycan array result of GST-tagged VP8*. The X-axis represents the CFG glycan number, and the Y-axis represents the relative fluorescence units (RFUs) of binding. Standard deviation from the mean of 6 replicates of each glycan printed on the array is provided ($n = 6$). **b** The glycan structures with the highest binding signals were drawn using GlycoGlyph⁴⁹. **c** Molecular docking of the LacNAc disaccharide (Gal-GlcNAc) onto VP8*B. The highest-ranked docked glycan is shown in the stick model, and the protein is shown with a transparent surface. The inset shows how the glycan interacts with VP8*B residues. The glycan and the VP8*B residues are shown in stick and ball-and-stick models, respectively.

Expression, purification, and crystallization of RVB VP8*. VP8*B (residues 65–233) of human rotavirus group B (GenBank: AET79992.1) was cloned into expression vectors pQE60 (Qiagen) with a C-terminal His tag and pGEX-4T-1 (Cytiva) with an N-terminal GST-tag. Recombinant C-terminal His-tagged VP8*B was expressed in *E. coli* BL21(DE3) cells (Novagen) by inducing cells with 200 μ M IPTG at 25 °C for 16 h. Cells were mechanically lysed and purified with Ni-NTA resin by batch purification. The His-tagged VP8*B was further purified by size exclusion column Superdex75 (GE healthcare) with 10 mM Tris, pH 8.0, 100 mM NaCl at 4 °C. The concentration of the purified protein was determined by measuring the absorbance at 280 nm and using an absorption coefficient of 12,950 $M^{-1} cm^{-1}$ for VP8*B calculated using ProtParam on the ExPASy server⁴².

Recombinant N-terminal GST-tagged VP8*B was expressed in *E. coli* BL21(DE3) cells (Novagen) by inducing cells with 200 μ M IPTG at 25 °C for 16 h. Cells were mechanically lysed and purified with Glutathione Sepharose resin by batch purification (Thermo Scientific). The purified elution was then dialyzed into 10 mM Tris pH 8.0, 100 mM NaCl overnight. The GST-VP8* was further purified by size exclusion column Superdex200 (GE healthcare) with 10 mM Tris pH 8.0, 100 mM NaCl at 4 °C. The concentration of the purified protein was determined by measuring the absorbance at 280 nm and using an absorption coefficient of 56,185 $M^{-1} cm^{-1}$ for GST-VP8*B calculated using ProtParam on the ExPASy server⁴².

Crystallization screenings for VP8*B at the concentration of 20 mg/ml were carried out by hanging-drop vapor diffusion using the Mosquito crystallization robot (TTP LabTech) at 20 °C. After 80 days, VP8*B was crystallized under the

condition with 0.2 M ammonium sulfate, 30% PEG 2 K MME, 0.1 M Na acetate, pH 4.6. Crystals were flash-frozen directly in liquid nitrogen.

Data collection and structure determination. X-ray diffraction data for the VP8*B crystals were collected on a PILATUS detector of beamline 5.0.1 at Advanced Light Source (Berkeley, CA). Diffraction data were processed using HKL2000⁴³. The flexible loops at the N- and C-termini of the highest-ranked AlphaFold2 model were removed, and only the residues 83–192 were used as the search model for MR using Phaser⁴⁴. A single solution with two molecules of VP8*B in the asymmetric unit was found by Phaser. The automated model building with the MR solution was carried out using ARP/wARP⁴⁴. Iterative cycles of refinement with simulated annealing and manual model building were performed using Phenix⁴⁵ and Coot³⁷. Data refinement and statistics are shown in Table 1. Figures were prepared using Chimera⁴⁶.

Glycan array screening. The glycan-binding specificity of VP8*B was investigated on a glycan array comprised of 600 glycans at the Emory Glycomics and Molecular Interactions Core (EGMIC), Emory University^{22–24}. Recombinant GST-tagged VP8*B at 5 μ g/ml concentration in binding buffer (20 mM Tris-HCl pH 7.4, 150 mM sodium chloride, 2 mM calcium chloride, 2 mM magnesium chloride, 0.05% Tween 20, 1% BSA) was applied to the glycan array, and bound protein was detected using 5 μ g/ml Alexa Flour 647 anti-GST tag monoclonal antibody (8–326) (Thermo Fisher Scientific, cat# MA4-004-A647). A summary of the glycan array

result is given in Supplementary Table 2. The complete list of glycans is provided in Supplementary Data 1.

Molecular docking with AutoDock Vina. The LacNAc disaccharide (Gal-GlcNAc) was docked onto VP8*B using AutoDock Vina⁴¹. The protein was processed by adding polar hydrogen atoms using AutoDockTools⁴⁷. VP8*B was treated as a rigid body, while the glycan was allowed to have all possible rotational angles. The grid box was centered on VP8*B with the size (40 × 40 × 40 Å) of the box adjusted to cover the entire protein. Docking was carried out with an exhaustiveness of 24. The pose with the lowest binding free energy of −5.7 kcal/mol was viewed using ViewDock in Chimera, and the molecular interactions were analyzed using LigPlot+ (v2.2.4)⁴⁸.

Evolutionary conservation analysis. Evolutionary conservation was analyzed using the ConSurf web server³⁹ using the full-length ADRV VP4B AlphaFold model as a query with default parameters. Multiple sequence alignment was built using MAFFT. The homologs were collected from non-redundant (NR) sequences from GenBank CDS translations with 35–95% identity. The calculation was performed with 30 unique sequences using the Bayesian calculation method.

Statistics and reproducibility. For the glycan array results, standard deviation and % coefficient of variation (%CV) are calculated from the mean of 6 replicates ($n = 6$) of each glycan printed on the array and are from one of two independent experiments with similar results.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Atomic coordinates and structure factors for the crystal structure of VP8*B have been deposited in the Protein Data Bank under the accession code 7RSW. The authors declare that all other data supporting the findings of this study are available within the paper and its supplementary information files. Source data are provided with this paper.

Received: 7 September 2021; Accepted: 12 April 2022;

Published online: 05 May 2022

References

- Estes, M. K. & Greenberg, H. B. in *Fields Virology* Vol. 2 (eds Knipe, D. M. & Howley, P. M.) Ch. 45, 1347–1401 (Lippincott Williams & Wilkins, 2013).
- Crawford, S. E. et al. Rotavirus infection. *Nat. Rev. Dis. Prim.* **3**, 17083 (2017).
- Matthijnsens, J. et al. VP6-sequence-based cutoff values as a criterion for rotavirus species demarcation. *Arch. Virol.* **157**, 1177–1182 (2012).
- Molinari, B. L., Lorenzetti, E., Otonel, R. A., Alfieri, A. F. & Alfieri, A. A. Species H rotavirus detected in piglets with diarrhea, Brazil, 2012. *Emerg. Infect. Dis.* **20**, 1019–1022 (2014).
- Bishop, R. F. Natural history of human rotavirus infection. *Arch. Virol. Suppl.* **12**, 119–128 (1996).
- Hung, T. et al. Rotavirus-like agent in adult non-bacterial diarrhoea in China. *Lancet* **2**, 1078–1079 (1983).
- Hung, T. et al. Waterborne outbreak of rotavirus diarrhoea in adults in China caused by a novel rotavirus. *Lancet* **1**, 1139–1142 (1984).
- Sanekata, T., Ahmed, M. U., Kader, A., Taniguchi, K. & Kobayashi, N. Human group B rotavirus infections cause severe diarrhea in children and adults in Bangladesh. *J. Clin. Microbiol.* **41**, 2187–2190 (2003).
- Krishnan, T. et al. Emergence of adult diarrhoea rotavirus in Calcutta, India. *Lancet* **353**, 380–381 (1999).
- Aung, T. S. et al. Detection of group B rotavirus in an adult with acute gastroenteritis in Yangon, Myanmar. *J. Med. Virol.* **81**, 1968–1974 (2009).
- Brown, D. W., Beards, G. M., Chen, G. M. & Flewett, T. H. Prevalence of antibody to group B (atypical) rotavirus in humans and animals. *J. Clin. Microbiol.* **25**, 316–319 (1987).
- Nakata, S. et al. Detection of antibody to group B adult diarrhoea rotaviruses in humans. *J. Clin. Microbiol.* **25**, 812–818 (1987).
- Uprety, T. et al. Identification of a ruminant origin group B rotavirus associated with diarrhea outbreaks in foals. *Viruses* **13**, <https://doi.org/10.3390/v13071330> (2021).
- Miyabe, F. M. et al. Porcine rotavirus B as primary causative agent of diarrhea outbreaks in newborn piglets. *Sci. Rep.* **10**, 22002 (2020).
- Viana, R. et al. Rapid epidemic expansion of the SARS-CoV-2 Omicron variant in southern Africa. *Nature* <https://doi.org/10.1038/s41586-022-04411-y> (2022).
- Fang, Z. Y. et al. Coding assignments of the genome of adult diarrhoea rotavirus. *Arch. Virol.* **125**, 53–69 (1992).
- Herrmann, T. et al. Functional refolding of the penetration protein on a non-enveloped virus. *Nature* **590**, 666–670 (2021).
- Settembre, E. C., Chen, J. Z., Dormitzer, P. R., Grigorieff, N. & Harrison, S. C. Atomic model of an infectious rotavirus particle. *EMBO J.* **30**, 408–416 (2011).
- Shaw, A. L. et al. Three-dimensional visualization of the rotavirus hemagglutinin structure. *Cell* **74**, 693–701 (1993).
- Banyai, K. et al. Candidate new rotavirus species in Schreiber's bats, Serbia. *Infect. Genet. Evol.* **48**, 19–26 (2017).
- Sun, X. et al. Human group C rotavirus VP8*s recognize type a histo-blood group antigens as ligands. *J. Virol.* **92**, <https://doi.org/10.1128/JVI.00442-18> (2018).
- Hu, L. et al. Structural basis of glycan specificity in neonate-specific bovine-human reassortant rotavirus. *Nat. Commun.* **6**, 8346 (2015).
- Hu, L. et al. Cell attachment protein VP8* of a human rotavirus specifically interacts with A-type histo-blood group antigen. *Nature* **485**, 256–259 (2012).
- Hu, L. et al. Glycan recognition in globally dominant human rotaviruses. *Nat. Commun.* **9**, 2631 (2018).
- Yu, X. et al. Novel structural insights into rotavirus recognition of ganglioside glycan receptors. *J. Mol. Biol.* **413**, 929–939 (2011).
- Xu, S. et al. Molecular basis of P[II] major human rotavirus VP8* domain recognition of histo-blood group antigens. *PLoS Pathog.* **16**, e1008386 (2020).
- Gozalbo-Rovira, R. et al. Unraveling the role of the secretor antigen in human rotavirus attachment to histo-blood group antigens. *PLoS Pathog.* **15**, e1007865 (2019).
- Dormitzer, P. R., Sun, Z. Y., Wagner, G. & Harrison, S. C. The rhesus rotavirus VP4 sialic acid binding domain has a galectin fold with a novel carbohydrate binding site. *EMBO J.* **21**, 885–897 (2002).
- Xu, S. et al. Structural basis of P[II] rotavirus evolution and host ranges under selection of histo-blood group antigens. *Proc. Natl Acad. Sci. USA* **118**, <https://doi.org/10.1073/pnas.2107963118> (2021).
- Ramani, S., Hu, L., Venkataram Prasad, B. V. & Estes, M. K. Diversity in rotavirus-host glycan interactions: a “Sweet” spectrum. *Cell Mol. Gastroenterol. Hepatol.* **2**, 263–273 (2016).
- Ramani, S. et al. The VP8* domain of neonatal rotavirus strain G10P[11] binds to type II precursor glycans. *J. Virol.* **87**, 7255–7264 (2013).
- Lahon, A. & Chitambar, S. D. Molecular characterization of VP4, VP6, VP7 and NSP4 genes of group B rotavirus strains from outbreaks of gastroenteritis. *Asian Pac. J. Trop. Med.* **4**, 846–849 (2011).
- Yang, J. et al. Improved protein structure prediction using predicted interresidue orientations. *Proc. Natl Acad. Sci. USA* **117**, 1496–1503 (2020).
- Yang, J. et al. The I-TASSER Suite: protein structure and function prediction. *Nat. Methods* **12**, 7–8 (2015).
- Tunyasyunakool, K. et al. Highly accurate protein structure prediction for the human proteome. *Nature* <https://doi.org/10.1038/s41586-021-03828-1> (2021).
- Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* <https://doi.org/10.1038/s41586-021-03819-2> (2021).
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr D. Biol. Crystallogr* **66**, 486–501 (2010).
- Holm, L. Using dali for protein structure comparison. *Methods Mol. Biol.* **2112**, 29–42 (2020).
- Ashkenazy, H. et al. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res.* **44**, W344–W350 (2016).
- Zheng, T. et al. Tracking N-acetylglucosamine on cell-surface glycans in vivo. *Angew. Chem. Int. Ed. Engl.* **50**, 4113–4118 (2011).
- Trott, O. & Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **31**, 455–461 (2010).
- Wilkins, M. R. et al. Protein identification and analysis tools in the ExPASy server. *Methods Mol. Biol.* **112**, 531–552 (1999).
- Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
- Winn, M. D. et al. Overview of the CCP4 suite and current developments. *Acta Crystallogr D. Biol. Crystallogr* **67**, 235–242 (2011).
- Liebschner, D. et al. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr D. Struct. Biol.* **75**, 861–877 (2019).
- Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
- Morris, G. M. et al. AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J. Comput. Chem.* **30**, 2785–2791 (2009).

48. Laskowski, R. A. & Swindells, M. B. LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. *J. Chem. Inf. Model* **51**, 2778–2786 (2011).
49. Mehta, A. Y. & Cummings, R. D. GlycoGlyph: a glycan visualizing, drawing and naming application. *Bioinformatics* **36**, 3613–3614 (2020).

Acknowledgements

We acknowledge the support from NIH grants AI36040 (to B.V.V.P.), AI080656 and P30 DK56338 (to M.K.E. and S.E.C.), and the Robert Welch Foundation (Q1279) to B.V.V.P. W.S. was supported through the training fellowship from the Gulf Coast Consortia, on the Training Interdisciplinary Pharmacology Scientists (TIPS) Program (Grant No. T32 GM120011). This research used the Advanced Light Source resources, a DOE Office of Science User Facility under contract no. DE-AC02-05CH11231. The ALS-ENABLE beamlines are supported by the National Institutes of Health, National Institute of General Medical Sciences, grant P30 GM124169-01.

Author contributions

L.H. and B.V.V.P. designed the research. L.H. predicted the AlphaFold2 model, determined the VP8*B crystal structure, and analyzed the data. W.S. performed the protein purification and crystallization, and B.S. performed X-ray diffraction data collection. Y.L. and D.F.S. contributed to glycan array experiments and analysis. S.E.C. and M.K.E. provided advice on the result analyses. L.H. and B.V.V.P. wrote the manuscript. All authors reviewed, edited, and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s42003-022-03357-1>.

Correspondence and requests for materials should be addressed to Liya Hu or B. V. Venkataram Prasad.

Peer review information *Communications Biology* thanks the anonymous reviewers for their contribution to the peer review of this work. Primary Handling Editors: Theam Soon Lim and Karli Montague-Cardoso. Peer reviewer reports are available.

Reprints and permission information is available at <http://www.nature.com/reprints>

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



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022