Structural basis of P[II] rotavirus evolution and host ranges under selection of histo-blood group antigens

Shenyuan Xu[a,b], Kristen Rose McGinnis[a], Yang Liu[c], Pengwei Huang[d], Ming Tan[d,e], Michael Robert Stuckert[a], Riley Erin Burnside[a], Elsa Grace Jacob[a], Shuisong Ni[a], Xi Jiang[d,e,1], and Michael A. Kennedy[a,1]

[a]Department of Chemistry and Biochemistry, Miami University, Oxford, OH 45056; [b]Key Laboratory of Bioorganic Synthesis of Zhejiang Province, College of Biotechnology and Bioengineering, Zhejiang University of Technology, Hangzhou 310014, China; [c]Tianjin Key Laboratory of Molecular Nuclear Medicine, Institute of Radiation Medicine, Chinese Academy of Medical Sciences, Peking Union Medical College, Tianjin 100730, China; [d]Division of Infectious Diseases, Cincinnati Children's Hospital Medical Center, Cincinnati, OH 45229; and [e]Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH 45267

Edited by Linda J. Saif, The Ohio State University, Wooster, OH, and approved August 2, 2021 (received for review April 27, 2021)

Group A rotaviruses cause severe gastroenteritis in infants and young children worldwide, with P[II] genogroup rotaviruses (RVs) responsible for >90% of global cases. RVs have diverse host ranges in different human and animal populations determined by host histo-blood group antigen (HBGA) receptor polymorphism, but details governing diversity, host ranges, and species barriers remain elusive. In this study, crystal structures of complexes of the major P[II] genogroup P[4] and P[8] genotype RV VP8* receptor–binding domains together with Lewis epitope–containing LNDFH I glycans in combination with VP8* receptor-glycan ligand affinity measurements based on NMR titration experiments revealed the structural basis for RV genotype-specific switching between ββ and βα HBGA receptor–binding sites that determine RV host ranges. The data support the hypothesis that P[II] RV evolution progressed from animals to humans under the selection of type 1 HBGAs guided by stepwise host synthesis of type 1 ABH and Lewis HBGAs. The results help explain disease burden, species barriers, epidemiology, and limited efficacy of current RV vaccines in developing countries. The structural data has the potential to impact the design of future vaccine strategies against RV gastroenteritis.

rotavirus | histo-blood group antigen | P[8] | LNDFH I | protein crystallography

The major human rotaviruses (RVs), the P[8], P[4], and P[6] genotypes in the P[II] genogroup, are responsible for over 90% of human infections worldwide (1–3). Despite successes of the RotaTeq and Rotarix RV vaccines in many developed countries, their efficacy remains disappointingly poor in developing countries (4–6). Low efficacies of both vaccines in developing countries can be attributed to a lack of cross protection between P[8], which is more common in developed countries, and other P-type RVs, such as P[6] and P[11], that are less common in developed countries but more common in developing countries (7–13).

Significant advances have been made in understanding RV evolution under the selection of stepwise synthesis of histo-blood group antigens (HBGAs) in humans. For example, P[II] RVs that mainly infect humans are thought to have originated from P[I] RVs with an animal host origin and evolved the ability to infect humans under selective pressure to bind polymorphic human HBGAs. This deduction is in agreement with a complete VP4 sequence phylogeny analysis that revealed that P[10]/P[12] in P[I] were genetically closer to P[19], P[6], and P[4]/P[8] in P[II] than other genotypes from other genogroups (14, 15). These observations led to the hypothesis that host ranges of P[II] genotypes for certain animal species and different human populations are dictated by the evolutionary stages of their HBGA receptors.

P[19] appears to represent an early evolutionary branch of the P[II] genogroup since it recognizes type 1 precursor HBGAs and therefore commonly infects animals (porcine) but rarely humans. On the other hand, P[4] and P[8] appear to be more evolutionarily advanced since they have developed the ability to recognize more mature HBGA products that dominate in humans. P[6] appears to represent an intermediate stage of evolution close to P[19] that commonly infects both animals (porcine) and humans, likely because of its evolutionary status that allows it to recognize less mature type 1 HBGA precursor glycans shared between humans and animals (porcine). The deduced evolutionary path that enabled the transition from animal to human host, which is correlated with the emergence of the P[II] branch from the P[I] branch, may apply to other genotypes and genogroups and may be important for RV classification and epidemiology (16, 17).

Evidence for HBGA-controlled RV host ranges and evolution is also available from structural analyses of genotype-specific interactions of RV VP8* domains with their glycan receptor ligands. For example, early structures showed that VP8* domains from animal and human RVs adopted similar galectin-like folds, and they recognize distinct HBGAs either through a ββ or βα site. However, our recent NMR spectroscopy–based docking and crystallographic studies showed that P[4], P[6], P[8], and P[19] VP8*s of P[II] interacted with H type 1 HBGA precursor using a common βα site (17–21), while P[8] VP8* bound Le^b tetra-saccharide and Lewis epitope–containing hexa-saccharide (LNDFH I) in the ββ site (22).

Significance

Group A rotaviruses cause severe gastroenteritis in infants and young children worldwide, with P[II] genogroup rotaviruses (RVs) responsible for >90% of global cases. Existing RV vaccines are successful in many developed countries, but their efficacy remains poor in developing countries, which could be due to lack of cross protection between RV genotypes more common in developed countries and other P-type RVs less common in developed countries but common in developing countries. In this study, crystal structures of major P[II] genogroup VP8* receptor domains in complex with Lewis epitope–containing histo-blood group antigen (HBGA) glycans were determined, revealing the structural basis for RV HBGA receptor specificities that determine RV host ranges. The data should enable rationale design of next-generation RV vaccines.

Author contributions: S.X., X.J., and M.A.K. designed research; S.X., K.R.M., Y.L., P.H., M.R.S., R.E.B., E.G.J., and S.N. performed research; S.X., M.T., X.J., and M.A.K. analyzed data; and S.X., M.T., X.J., and M.A.K. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: kennedm4@miamioh.edu or jason.jiang@cchmc.org.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2107963118/-/DCSupplemental.

Published September 2, 2021.

To elucidate the molecular basis for receptor-binding bias between βα- and ββ-binding sites, we characterized relative binding affinities of major P[II] RV VP8* domains for glycans representing different HBGA synthetic stages, including the Lewis epitope–containing LNDFH I, using NMR heteronuclear single quantum coherence spectroscopy (HSQC)-monitored titrations. The structural basis for the bias of ββ sites for Lewis epitope HBGAs and βα sites for HBGAs lacking the Lewis epitope was elucidated from crystal structures of P[4] and P[8] bound to LNDFH I and P[6] bound to Lacto-*N*-tetraose (LNT). Sequence- and structure-based analyses of differences in P[II] VP8* receptor–binding interfaces revealed molecular details responsible for receptor switching between genotype-specific ββ and βα HBGA–binding sites. Overall, the results provide strong evidence for HBGA-controlled P[II] RV evolution from an animal host origin that resulted in diverse genotypes infecting children in different populations and which may impact future strategies for RV disease control and prevention.

Results

Crystal Structures of Apo P[4] and P[8] VP8* Domains. VP8* apoform crystal structures of human P[4] strain BM5265 and P[8] strain BM13851 were solved at 1.21 and 1.71 Å, respectively (Fig. 1*A* and Table 1). The VP8* domains adopted a classic galectin-like fold, with two twisted antiparallel β-sheets consisting of strands A (72 to 74), L (197 to 200), C (90 to 96), D (102 to 108), G (137 to 144), H (152 to 159) and M (204 to 208), B (80 to 85), I (163 to 169), J (172 to 177), K (184 to 189), respectively. Other secondary structural elements included βE (115 to 121), βF (124 to 130), and αA (212 to 221). P[4] and P[8] VP8* shared similar overall conformations, with the root-mean-squared deviation (RMSD) for backbone alpha carbons being 0.641 Å. The distinct cleft formed by two β-sheets was noticeably wider in P[8] than in P[4] VP8*, and the I–J loop conformation was slightly different between P[4] and P[8] VP8* (Fig. 1*A*), which may impact glycan specificity and affinity.

Relative Binding Affinities of P[II] VP8*s for Lewis B Positive and Negative HBGA Glycans Using NMR HSQC Titration Experiments. Three type 1 HBGA glycans were used in NMR HSQC titration experiments: LNT (Galβ1–3GlcNAcβ1–3Galβ1–4Glc) that contained the type I precursor but lacked both Secretor and Lewis fucoses, Lacto-*N*-fucopentaose I (LNFP I: Fucα1–2Galβ1–3GlcNAcβ1–3Galβ1–4Glc) that contained the Secretor fucose (Fucα1–2), and Lacto-*N*-difucohexaose I (LNDFH I: Fucα1–2Galβ1–3[Fucα1–4]GlcNAcβ1–3Galβ1–4Glc) that contained both Secretor and Lewis fucoses (Fucα1–4) (Fig. 1*B*). These representative HBGAs enabled the investigation of roles that Secretor fucose and Lewis fucose played in determining the binding specificity of P[II] RVs at different evolutionary stages to HBGAs at different biosynthetic stages. P[4] VP8* backbone resonances were 91% complete. Backbone assignments of P[6] and P[19] were reported previously (17, 22).

The titration of P[4] VP8* with LNDFH I caused a disappearance and chemical shift perturbation of resonances belonging to amino acids in the ββ site (*SI Appendix*, Fig. S1*A*), indicating P[4] binds HBGA glycans containing both Lewis fucose and Secretor fucose using the ββ site (Fig. 2*A*). Affected residues included F139 and E141 of βG; R154, R155, T156, and L157 of βH; H177, G178, and E179 of βJ; A183 of the βJ–βK loop; and T184, T185, and D186 of βK plus T115, N132, and S134 away from the putative ββ-binding site (Fig. 2 *B* and *C*). The dissociation constant, K_d, for P[4] binding LNDFH I was 2.5 ± 0.2 mM (*SI Appendix*, Fig. S2*A*).

The titration of P[4] VP8* with LNFP I caused chemical shifts for amino acids in the βα site (*SI Appendix*, Fig. S3*A*), indicating P[4] VP8* binds HBGA glycans lacking the Lewis fucose in the βα site (*SI Appendix*, Fig. S4 *A*–*C*). These residues included Y169 of βI, F176 of βJ, T185 and D186 of βK, R209 of the βM–αA loop, and E212, C215, and E217 of αA plus K138 and G145 away from the βα site. The K_d for P[4] VP8* binding LNFP I was 12.5 ± 4.6 mM (*SI Appendix*, Fig. S5*A*), that is, fivefold weaker compared to LNDFH I. P[4] binding to LNT was not detected (*SI Appendix*, Fig. S3*C*).

The titration of P[6] VP8* with LNDFH I caused chemical shifts for amino acids in the ββ site (*SI Appendix*, Fig. S1*B*), indicating P[6] binds HBGA glycans containing the Lewis fucose in the ββ site (Fig. 2 *D*–*F*). The residues involved included K154,

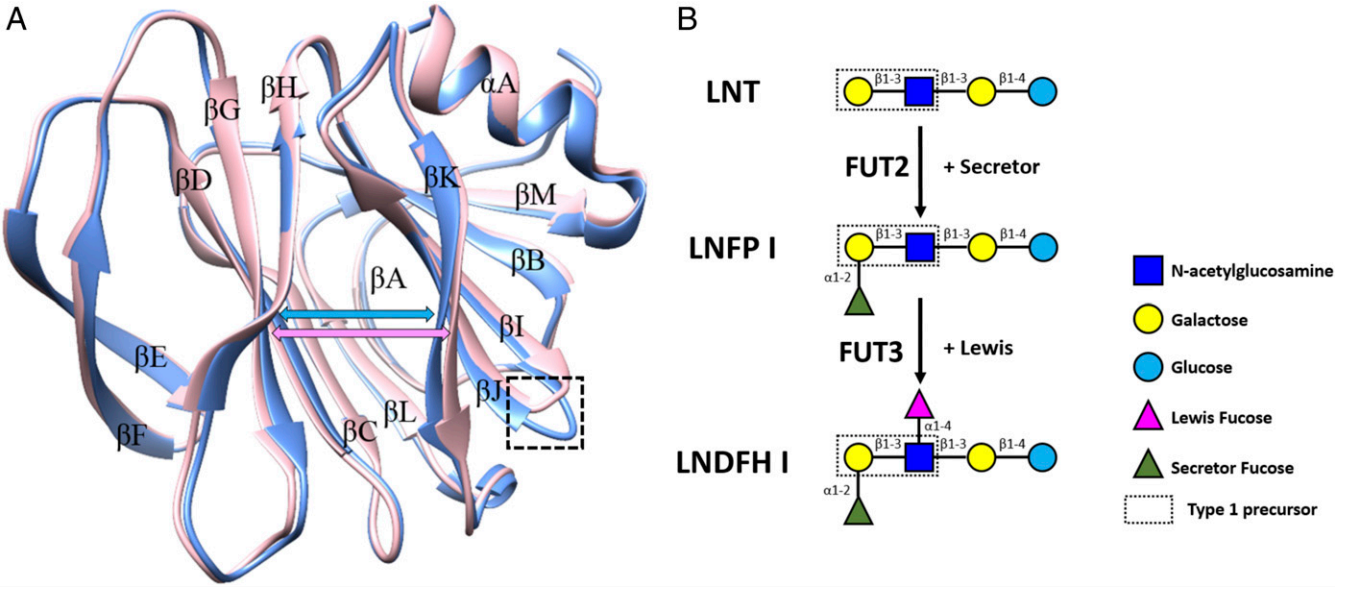

Fig. 1. Crystal structures of apo P[4]/P[8] human RV VP8* domains and glycan structures. (*A*) Ribbon representation of P[4] (blue, PDB ID: 6UT9) human BM5265 and P[8] (pink, PDB ID: 6VKX) human BM13851 VP8* with their structures superposed. Secondary structures are labeled with letters. The blue double arrow and pink double arrow represent the width of the cleft between the two β sheets in the P[4] and P[8] VP8*, respectively. The square dashed box highlights the difference in the I–J loop. (*B*) Schematic representation of glycan structures used for binding titrations in this study. Glycan reagents were LNT (Lacto-*N*-tetraose), LNFP I (Lacto-*N*-fucopentaose I), and LNDFH I (Lacto-*N*-difucohexaose I). FUT2 represents the α1,2-fucosyltransferase (Secretor enzyme), and FUT3 represents the α1,3/4-fucosyltransferase (Lewis enzyme).

Table 1. Diffraction data and structure refinement statistics

	P[8] BM13851 VP8*	P[4] BM5265 VP8*	P[8] BM13851 VP8* -LNDFH I	P[4] BM5265 VP8- LNDFH I	P[6] BM11596 VP8*- LNT
PDB ID	6VKX	6UT9	7JHZ	7KHU	7KI5
Crystal parameters					
Space group	C 1 2 1	P 1 21 1	C 1 2 1	P 1	P 1 21 1
Unit cell parameters					
a; b; c (Å)	112.00; 63.25; 71.94	47.04; 32.45; 51.00	112.19; 63.39; 72.26	36.32; 38.43; 55.81	56.68; 75.35; 74.41
α; β; γ (°)	90.00; 90.70; 90.00	90.00; 94.64; 90.00	90.00; 91.15; 90.00	104.73; 90.15; 96.04	90.00; 91.77; 90.00
Data collection					
Wavelength (Å)	0.97931	0.97931	0.97931	0.97932	0.97932
Resolution (Å)	56.00–1.71 (1.80–1.71)	50.83–1.21 (1.27–1.21)	56.08–2.68 (2.82–2.68)	53.95–2.54 (2.58–2.54)	74.38–1.52 (1.60–1.52)
R_{merge}	0.049 (0.410)*	0.048 (0.383)	0.170 (0.598)	0.195(0.455)	0.051 (0.578)
CC(1/2)	0.998 (0.799)	0.998 (0.850)	0.977 (0.764)	0.979 (0.840)	0.998 (0.804)
Unique reflections	48,913 (7,683)	44,980 (6,692)	13,841 (2,043)	9,452 (484)	92,741 (13,335)
Mean [(I)/σ(I)]	12.7 (2.4)	10.5 (2.5)	7.1 (2.5)	5.9 (1.7)	12.4 (2.6)
Completeness	90.0 (97.4)	95.1 (97.6)	95.9 (97.6)	98.8 (99.8)	96.1 (94.9)
Multiplicity	3.3 (3.4)	3.3 (3.3)	5.1 (5.3)	6.6 (6.3)	3.4 (3.4)
Wilson B factor	19.9	9.2	17.6	21.7	15.2
Refinement					
Resolution (Å)	32.17–1.71 (1.80–1.71)	50.83–1.21 (1.27–1.21)	56.08–2.68 (2.82–2.68)	53.95–2.54 (2.58–2.54)	74.38–1.52 (1.60–1.52)
R-work	0.198	0.187	0.211	0.208	0.175
R-Free	0.228	0.211	0.263	0.260	0.196
Number of protein atoms	3,812	1,297	3,802	2,635	5,539
Number of amino acid residues	469	161	449	317	631
Number of ligands	3 (1 PG4, 2 PEG)	0	4 (2 LNDFH I, 2 GOL)	2 (2 LNDFH I)	3 (1 LNT, 1 PGE, 1 GOL)
Number of water molecules	117	138	9	45	281
Mean B-values					
Protein	23.74	12.49	25.28	28.29	21.02
Ligands	39.02	0	34.48	25.67	36.05
Solvent	24.64	20.84	15.06	19.87	29.11
RMS deviation					
Bond lengths (Å)	0.013	0.016	0.010	0.010	0.014
Bond angles (°)	1.657	1.470	1.593	1.689	1.233
Ramachandran statistics (%)					
Preferred regions	95.66	97.48	96.57	94.93	96.31
Allowed regions	4.34	2.52	3.43	5.07	3.69
Outliers	0	0	0	0	0

*Values in parentheses are for the highest-resolution shell.

R155, and T156 of βH; G178 and E179 of βJ; A183 of the βJ–βK loop; T184, T185, D186, and Y187 of βK; and residue K168 away from the ββ site. The K_d for P[6] binding LNDFH I was 30.4 ± 13.2 mM (*SI Appendix*, Fig. S2*B*), that is, ~12-fold weaker binding compared to P[4] VP8*.

The titration of P[19] VP8* with LNDFH I caused chemical shifts for residues in the ββ site (*SI Appendix*, Fig. S1*C*), indicating P[19] VP8* binds HBGA glycans containing the Lewis fucose in the ββ site (Fig. 2 *G–I*). The affected residues included T156 and T158 of βH, G178 and E179 of βJ, A183 of the βJ–βK loop, and T184, T185, D186, and Y187 of βK. The K_d for P[19] binding LNDFH I was 48.2 ± 38.9 mM (*SI Appendix*, Fig. S2*C*), indicating nearly 20-fold weaker binding compared to P[4] VP8*.

The titration of P[19] with LNT caused chemical shifts for residues in the βα site (*SI Appendix*, Fig. S4 *D–F*), indicating P[19] binds HBGA glycans lacking the Lewis fucose in the βα site (*SI Appendix*, Fig. S3*B*). The affected residues included D79 of the A–B loop; Y80 of βB; K168 and H169 of βI; R172, L173, and W174 of βJ; T184 and T185 of βK; R209 of the βM–αA loop; E212, K214, C215, T216, and E217 of αA; and for Y201 away from the βα site. The K_d for P[19] binding LNT was 6.3 ± 0.6 mM (*SI Appendix*, Fig. S5*B*).

Molecular Basis for P[8] VP8* Recognition of LNDFH I. The crystal structure of P[8] VP8* in complex with LNDFH I was solved at 2.68 Å (Table 1). There were three chains in the asymmetric unit, and the electron density for LNDFH I was clear in chains A and C (except the glucose moiety in chain A) (*SI Appendix*, Fig. S6 *A–D*). Ligand binding did not significantly perturb the protein structure (backbone Cα RMSD between apo and complex = 0.230 Å). LNDFH I inserted into the ββ site and occupied nearly the entire shallow cleft (Fig. 3*A*), binding through a network of hydrogen-bonding and hydrophobic interactions (Fig. 3 *B* and *C*). The galactose of the type 1 precursor in LNDFH I (GAL-II) did not make direct contact with the P[8] VP8* surface but appeared to stabilize LNDFH I in its bound conformation for interaction with P[8] VP8*. *N*-acetylglucosamine of the type 1 precursor (GlcNAc-III) was situated in the center of the ββ site and formed hydrophobic interactions with H177 and hydrogen bonds with T156. The Secretor fucose (Fuc-I) and Lewis fucose (Fuc-IV) served as anchor points in attaching LNDFH I to the

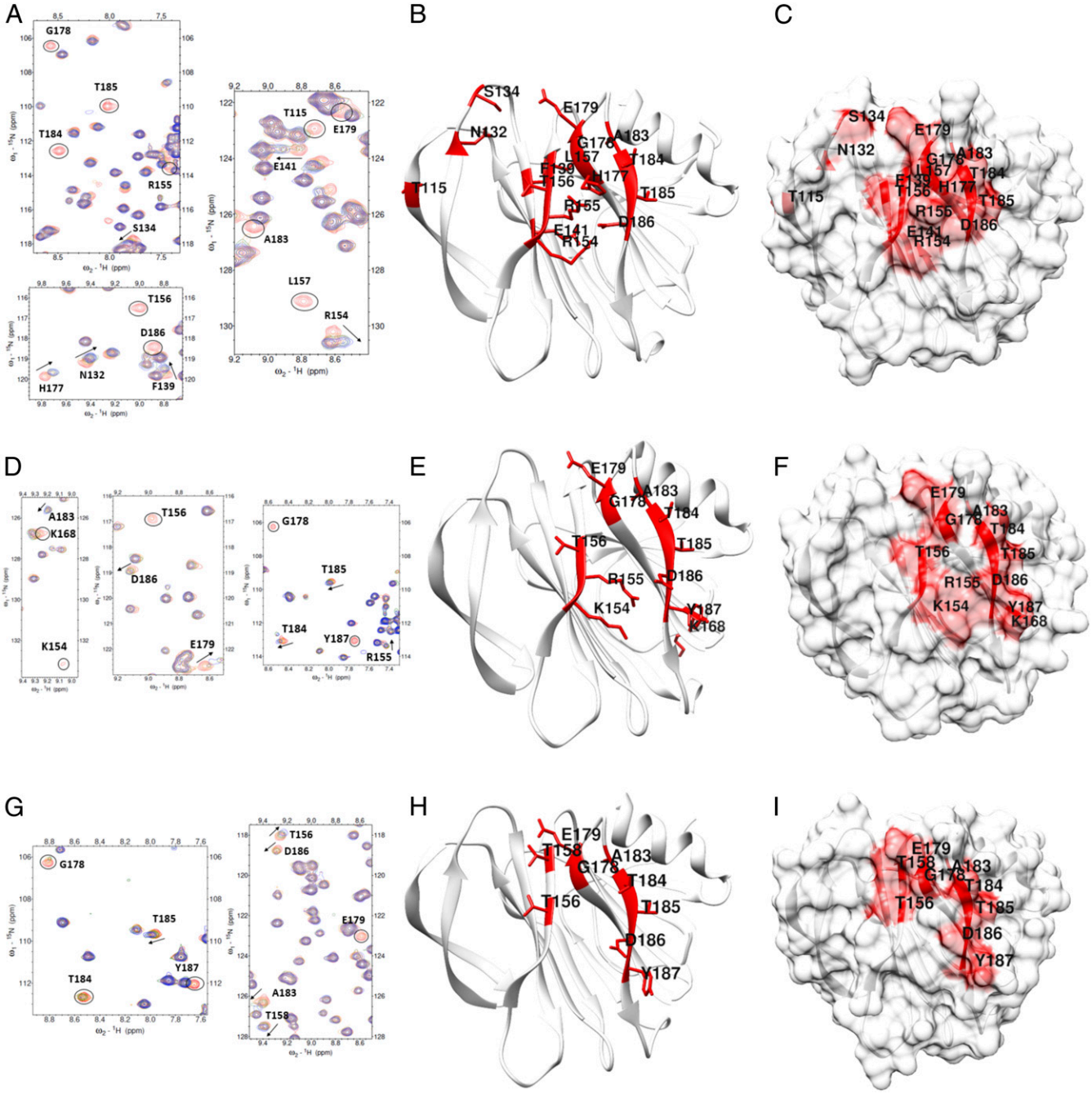

Fig. 2. Titrations of P[4], P[6], and P[19] VP8* with LNDFH I. (*A*) Representative spectra showing chemical shift changes of P[4] VP8* upon addition of LNDFH I. (*B*) Ribbon diagram indicating residues that disappeared or had large chemical shift changes on P[4] VP8* (PDB ID: 6VKX) and (*C*) surface diagram showing the binding interface of P[4] VP8* with LNDFH I. (*D*) Representative spectra showing chemical shift changes of P[6] VP8* upon addition of LNDFH I. (*E*) Ribbon diagram indicating residues that disappeared or had large chemical shift changes on P[6] VP8* (PDB ID: 6NIW) and (*F*) surface diagram showing the binding interface of P[6] VP8* with LNDFH I. (*G*) Representative spectra showing chemical shift changes of P[19] VP8* upon addition of LNDFH I. (*H*) Ribbon diagram indicating residues that disappeared or had large chemical shift changes on P[19] VP8* (PDB ID: 5VKS) and (*I*) surface diagram showing the binding interface of P[19] VP8* with LNDFH I. NMR titration data correspond to increasing ligand/protein ratios of 0:1 (red), 8:1 (orange), 16:1 (green), and 25:1 (blue). Residues involved in LNDFH I binding are colored red and labeled.

ββ site. The Secretor fucose occupied the pocket formed by the G178, E179, R182, and T184 and formed a hydrogen bond with T184. The Lewis fucose inserted into the pocket formed by T156, T158, G178, and E179. The fourth and fifth saccharide of LNDFH I, Gal-IV(β1,4)Glc-V, made direct contacts with the P[8] VP8* surface and contributed to the stabilization of the interaction. Gal-IV(β1,4)Glc-V interacted with Y152, R154, R155, S187, and S188 via hydrophobic interactions and formed hydrogen bonds with Y152, R154, R155, H177, D186, and S187.

Molecular Basis for P[4] VP8* Binding to LNDFH I. The crystal structure of LNDFH I–bound P[4] VP8* was solved at 2.54 Å (Table 1). The electron density was clear for the bound ligands in both chains in the asymmetric unit (*SI Appendix*, Fig. S6 *E–H*). Conformations of the apo and LNDFH I–bound P[4] VP8* were similar (backbone Cα RMSD = 0.646 Å). P[4] VP8* used the same shallow cleft of the ββ site as P[8] VP8* to bind LNDFH I (Fig. 4 *A–C*). Similar to the LNDFH I-P[8] VP8* complex, the binding of LNDFH I to P[4] was mediated through a network of

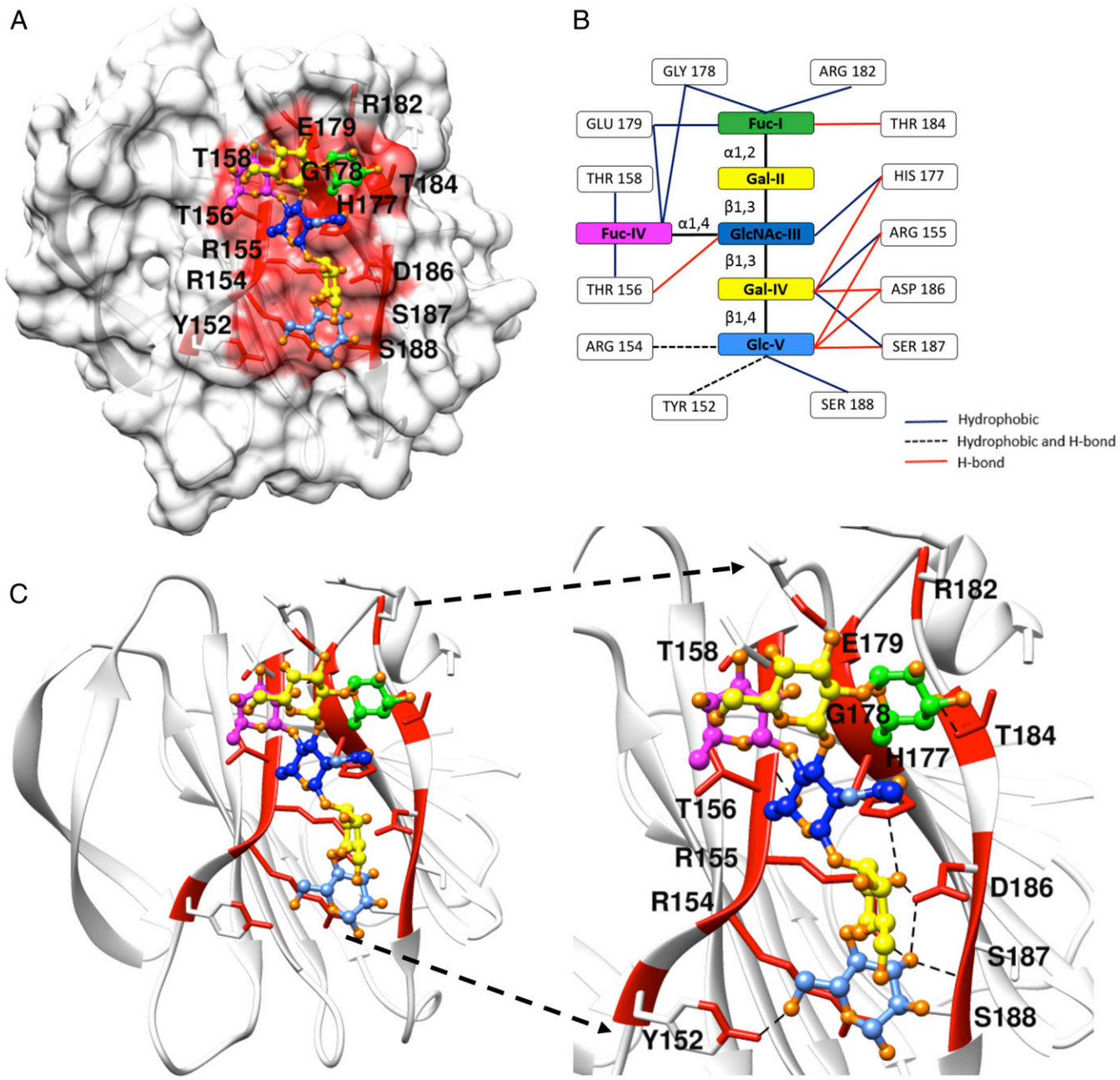

Fig. 3. Crystal structure of the LNDFH I-P[8] VP8* complex. (*A*) Surface model of LNDFH I-P[8] VP8* complex (PDB: 7JHZ) showing the recognition interface (colored red) between P[8] VP8* and LNDFH I. (*B*) Schematic diagram shows the interaction of P[8] VP8* with LNDFH I. (*C*) Ribbon diagram showing the LNDFH I-P[8] VP8* complex. (*C*, *Right*) Highlights the LNDFH I–binding pocket of P[8] VP8*. Different carbohydrate moieties of the glycan ligand were colored as follows: Lewis fucose (magenta), secretor fucose (green), galactose (yellow), *N*-acetylglucosamine (blue), and glucose (cornflower blue).

hydrogen-bonding and hydrophobic interactions (Fig. 4 *B* and *C*), with the central *N*-acetylglucosamine (GlcNAc-III), Secretor fucose (Fuc-I), and Lewis fucose (Fuc-IV) serving as building blocks to contact the binding interface. *N*-acetylglucosamine of the type 1 precursor (GlcNAc-III) formed hydrogen bonds with T156 and hydrophobic interactions with H177. The Secretor fucose and Lewis fucose occupied the pocket formed by the K138, T156, T158, G178, E179, R182, and T184, and the Secretor fucose formed a hydrogen bond with T184. The fourth saccharide of LNDFH I, GAL-IV, made direct contacts with the P[8] VP8* surface and formed hydrophobic interactions with R154 and R155 and hydrogen bonds with R154 and D186. The galactose of the type 1 HBGA precursor, Gal-II, made no contact with the binding site, and the glucose of LNDFH I pointed away from the binding pocket.

Molecular Basis for P[6] VP8* Recognition of LNT. The crystal structure of LNT-bound P[6] VP8* was solved at 1.52 Å (Table 1). While the electron density for the LNT ligand was weak due to weak binding affinity, the electron density around the βα-site adequately defined the location of the LNT ligand, which was supported by an OMIT map (*SI Appendix*, Fig. S6 *I* and *J*). The binding of LNT to P[6] VP8* did not perturb the P[6] VP8* structure (backbone Cα RMSD = 0.136 Å). P[6] VP8* bound LNT in the βα site through a network of hydrogen-bonding and hydrophobic interactions (Fig. 5). The type 1 precursor, Gal-II(β1,3)

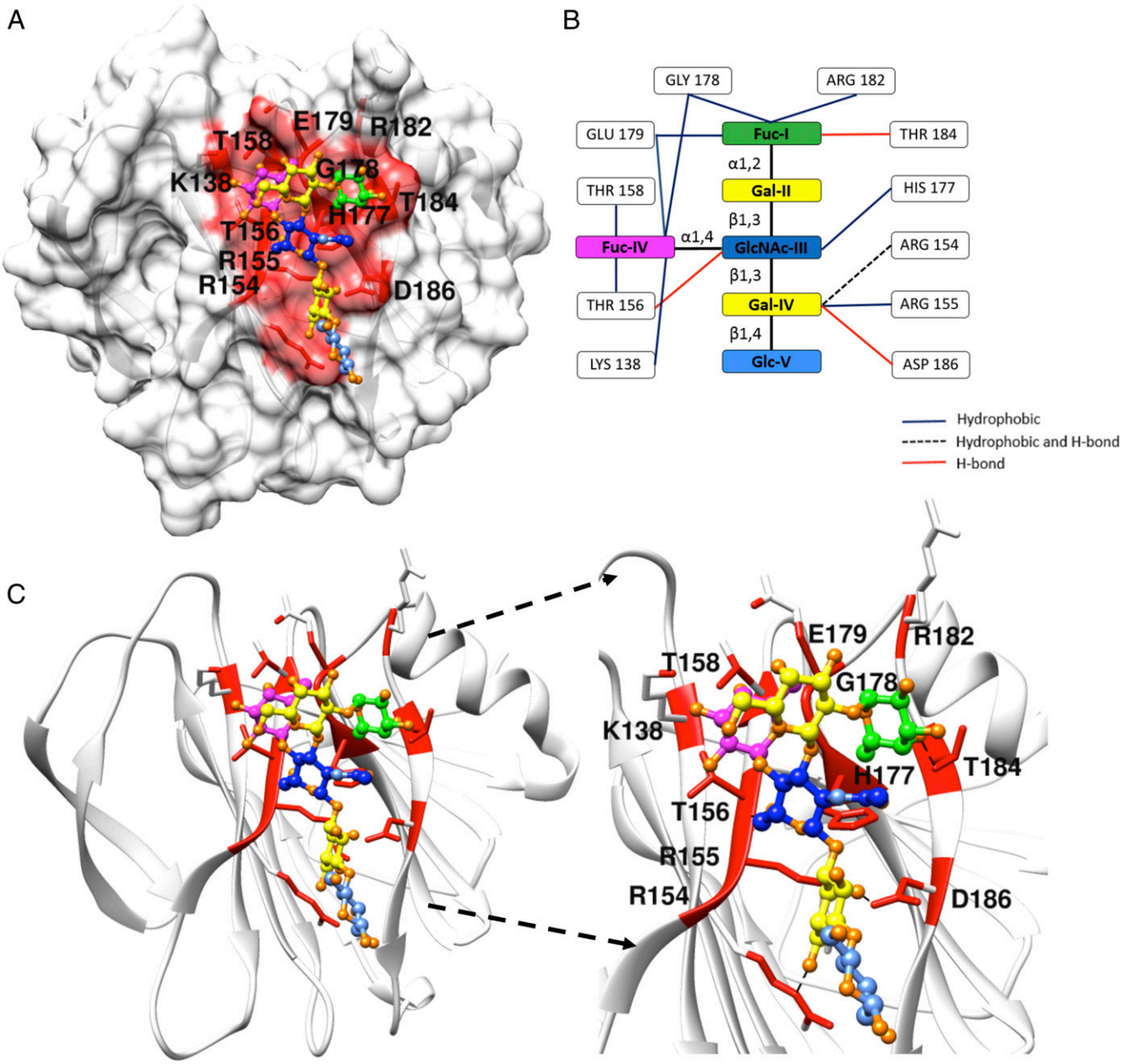

Fig. 4. Crystal structure of the LNDFH I-P[4] VP8* complex. (*A*) Surface model of the LNDFH I-P[4] VP8* complex (PDB: 7KHU) showing the interface of P[4] VP8* recognizing LNDFH I (colored red). (*B*) Schematic diagram showing the interaction of P[4] VP8* with LNDFH I. (*C*) Ribbon diagram showing the LNDFH I-P[4] VP8* complex. (*C*, *Right*) Highlights the binding pocket of P[4] VP8* in recognizing LNDFH I. Different carbohydrate moieties of the glycan ligand were colored as follows: Lewis fucose (magenta), secretor fucose (green), galactose (yellow), *N*-acetylglucosamine (blue), and glucose (cornflower blue).

GlcNAc-III, of LNT inserted into the pocket formed by W81, M167, W174, F176, A183, T184, T185, R209, and E212 with Gal-II forming hydrogen bonds with T184 and GlcNAc-III forming hydrogen bonds with T185, R209, and E212. The terminal galactose (Gal-IV) and glucose (Glc-V) of LNT lay in the pocket formed by F169, Y170, N171, S172, Y187, S189, and R209, and the glucose of LNT formed hydrogen bonds with Y170 and S172.

VP8* Amino Acid Positions that Diverged during P[II] RV Evolution. Sequence-based principal component analysis (PCA) (23) was performed on 85 sequences representing different genotypes of the P[II] genogroup to identify amino acid positions that varied most over the course of P[II] RV evolution. P[4], P[6], P[8], and P[19] genotypes were well separated in the eigenspace defined by PC1, PC2, and PC3 (*SI Appendix*, Fig. S7*A*). Residues that contributed the most variance were located away from the HBGA glycan ββ- and βα-binding sites, including 89, 113, 114, 115, 120, 131, 133, 140, and 166, and most amino acids involved in the binding interface were highly conserved, including 155, 156, 157, 158, 177, 178, 183, 185, 186, and 188 in the ββ-binding site and 81, 174, 185, 209, and 212 in the βα-binding site (*SI Appendix*, Fig. S7*B*). However, residues 153, 154, 179, 184, and 187 in the ββ-binding site and 167, 170, 171, 172, and 184 in the βα-binding site contributed relatively high variance, indicating that sequence variation outside and within the binding pockets contributed to RV evolution, resulting in VP8* binding HBGA glycans in a genotype-specific manner.

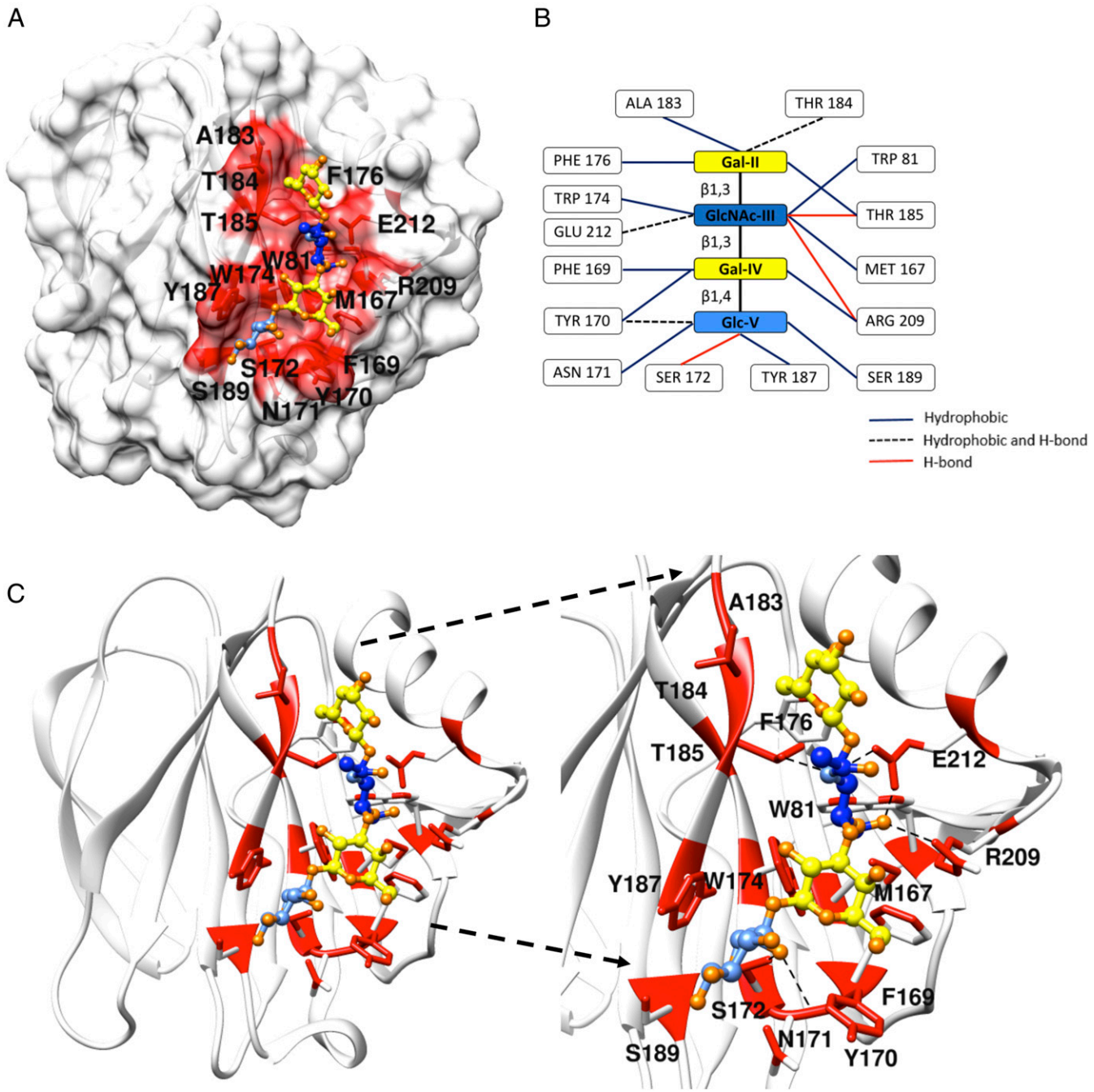

Fig. 5. Crystal structure of the LNT-P[6] VP8* complex. (*A*) Surface model of the LNT-P[6] VP8* complex (PDB: 7KI5) showing the interface of P[6] in recognizing LNT (colored red). (*B*) Schematic diagram shows the interaction of P[6] VP8* with LNT. (*C*) Ribbon diagram showing the LNT-P[6] VP8* complex. (*C, Right*) Highlights the binding pocket of P[6] VP8* in recognizing LNT. Different carbohydrate moiety of the glycan ligand were colored as follows: galactose (yellow), *N*-acetylglucosamine (blue), and glucose (cornflower blue).

Amino Acid Conservation in the Substrate-Binding Site of the P[II] RVs. Positional amino acid conservation in P[II] VP8* sequences was calculated using ConSurf (24) and mapped onto the structures (*SI Appendix*, Fig. S7*C*). Residues in the ββ- and βα-binding sites were most conserved, especially residues 155, 156, 158, 177, 186, and 188 in the ββ site and residues 168, 185, 209, and 212 in the βα site (*SI Appendix*, Fig. S7*D*). A structure-based sequence alignment of P[4], P[6], P[8], and P[19] with known glycan-bound structures also showed significant amino acid conservation in the ββ and βα glycan–binding interfaces (*SI Appendix*, Fig. S8*A*), consistent with phylogenetic tree results showing that they belong to the same P[II] genogroup with some shared glycan ligands, such as H type 1 HBGA.

Variations in Surface Electrostatic Potential in Substrate-Binding Sites of P[II] RVs. Surface electrostatic potential near residues 182, 184, 186, 187, and 188 of the ββ interface was significantly more negative in P[6] than in P[19] and P[4]/P[8] (*SI Appendix*, Fig. S8 *B–E*) and significantly more positive near residues 152 and 188 in P[19] than in P[6] and P[4]/P[8] (*SI Appendix*, Fig. S8 *B–E*). Electrostatic potential near residue 172 of the βα interface was more negative in P[6] than in P[19] and P[4]/P[8] (*SI Appendix*, Fig. S8 *F–I*), and

electrostatic potential near residues 81, 167, 174, 184, 185, 212, and 216 was more negative in P[6]/P[19] than in P[4]/P[8] (*SI Appendix*, Fig. S8 *F–I*). The implications of these differences between RV genotypes regarding ligand-binding specificity and affinity are discussed below in *Structure-Based Analysis of Relative VP8* HBGA Glycan–Binding Affinities*.

Pinpointing Amino Acid Changes Responsible for Divergence of P[II] RVs into Sub-P[II] Lineages (Genotypes) with Distinct ββ- and βα-Binding Sites for Lewis and Non-Lewis HBGAs. All structural data for VP8* domains binding to LNT, LNFP I, and LNDFH I ligands is summarized in Fig. 6. The superposition of LNDFH I onto the βα site in P[4] caused steric clashes with F176, T185, and E212 (*SI Appendix*, Fig. S9*K*), and the superposition of LNDFH I onto the βα site in P[8] caused clashes with W81, R172, F176, T185, and E212 (*SI Appendix*, Fig. S9*O*). These steric clashes caused an overestimation of contact surface areas and the number of hydrogen bonds (*SI Appendix*, Tables S1 and S3) and increased binding energy (*SI Appendix*, Table S2), explaining why LNDFH I in particular, and why Lewis epitope–containing HBGAs in general, do not bind in the VP8* βα site.

To investigate the structural basis for this bias, LNFP I structures bound in the βα site were superposed onto the common structural fragment within LNDFH I bound in the ββ site, and the resulting binding energies, binding surface areas, and hydrogen bonds were evaluated (*SI Appendix*, Fig. S9). When LNFP I was superposed into the ββ site of P[II], the binding energy increased (*SI Appendix*, Table S2), contact surface area decreased (*SI Appendix*, Table S3), and hydrogen bonds decreased (*SI Appendix*, Table S1) compared to when LNFP I bound to the βα site, indicating these energetic consequences do not favor LNFP I binding in the ββ site.

Structure-Based Analysis of Relative VP8* HBGA Glycan–Binding Affinities. Based on K_d values, P[6]/P[19] have higher affinity for LNT compared to P[4]/P[8] for which binding was not detected. When apo forms of P[4], P[8], and the LNT-P[6] VP8* crystal structures were superposed onto the crystal structure of P[19]-LNT (25) (*SI Appendix*, Fig. S10 *A–C*), it was apparent that H169 in P[19] was substituted with Y169 in P[4]/P[8], and this substitution must be responsible for the loss of P[4]/P[8] binding to LNT since all other residues in the binding interface were conserved. It appeared that the tyrosyl sidechain of Y169 sticks into the binding pocket, occluding LNT binding in P[4]/P[8]. The electrostatic surface potential near residue 169 switched from neutral or slightly negative in P[19] and P[6] to strongly positive in P[4] and P[8], which could contribute to the loss of affinity (*SI Appendix*, Fig. S8 *F–I*). The LNT orientation was different in the P[6] and P[19] complexes, particularly in the terminal galactose and glucose, due to changes in key residues involved in binding, for example, M167 in P[6] compared to L167 in P[19], F169 in P[6] compared to H169 in P[19], and Y170 in P[6] compared to G170 in P[19], and these changes appear to perturb the orientations of GlcNAc, Gal, and Glc, respectively, in P[6] in comparison to in P[19], leading to a complete reorganization of LNT binding in the βα-binding site (Fig. 6 *B–C*).

The binding affinity of VP8* to LNFP I decreased in the order P[19], P[6]/P[4], and P[8] (Table 2). Alternative conformations of LNFP I were observed in superpositions of P[4]-LNFP I, P[6]-LNFP I, P[8]-LNFP I, and P[19]-LNFP I (18, 19, 21, 22), especially in the Gal-IV(β1,4)Glc-V moiety (Fig. 6 *D–G* and *SI Appendix*, Fig. S10 *D–I*). The most noticeable difference was that substitutions of G170, G171, and R172 in P[19], P[4], and P[8] to Y170, N171, and S172 in P[6] pushed the glucose moiety of LNFP I (Glc-V) to the side of the βK strand of P[6] (Fig. 6*E*). The conformation of LNT and LNFP I in P[19] (Fig. 6 *B* and *D*, respectively) were almost the same with the extra Secretor fucose pointing away from the protein surface (*SI Appendix*, Fig. S10*J*), similar to the conformations observed between LNT and LNFP I in P[6] (*SI Appendix*, Fig. S10*K*). The Secretor fucose in LNFP I may stabilize the conformation required for P[4]/P[8] binding, albeit with a much weaker affinity than P[19].

P[4]/P[8] bound LNDFH I much more tightly than P[6]/P[19] (Table 2). Most residues in the ββ site of P[II] VP8* were similar except residues 152, 154, 182, and 187 (*SI Appendix*, Fig. S10 *L–N*). While there were some differences in electrostatic potentials of the protein surface among P[II] RVs, the substitution of R154 in P[4]/P[8] to K154 in P[6]/P[19] appeared to be responsible for a much weaker binding of P[6]/P[19] to LNDFH I compared to P[4]/P[8] due to a loss of hydrogen bonding between NH2 of R154 with O5 of the glucose moiety in LNDFH I–bound P[8] and between NE of R154 with O6 of the galactose moiety in LNDFH I–bound P[4] (Fig. 6 *H–I* and *SI Appendix*, Fig. S10 *L–N*). The terminal Gal-IV(β1,4)Glc-V of LNDFH I had different conformations when bound to P[4] and P[8], but binding to the Lewis fucose was established by hydrophobic interactions in both cases (Figs. 3 and 4).

Discussion

Crystal structures of P[4] and P[8] VP8*s in complex with LNDFH I reported here in combination with NMR analyses of additional P[II] RV genotypes and sequence-based analyses solidify the hypothesis that major P[II] human RV pathogens responsible for over 90% of global RV infections evolved from P[I] genogroup RVs with an animal host origin under a strong selection of type 1 HBGAs. We propose that ancestral viruses in the P[I] genogroup, such as P[10]/P[12], may have originated with a small receptor-binding interface that accommodated simple saccharide ligands, such as the disaccharide H type 1 precursor GlcNAc-Gal, and circulated in different animal species before being introduced to humans, which resulted in a continually expanding binding interface to accommodate more complicated H type 1 antigens generated by the stepwise addition of the A, B, H, and Lewis epitopes in the ABH and Lewis HBGA families, which are polymorphic in humans with different distributions across the world's population. For example, H-positive individuals who contain secretor antigens account for about 80% of European and North American populations (26), and Lewis-positive phenotypes occur in about 90% of the general population but with lower frequencies in Africa (27). Thus, the elucidation of species-specific HBGAs could be significant for the determination of host ranges and epidemiology, the mechanism of species barriers, and cross-species transmission.

P[19] appears to represent the earliest evolutionary branch in the P[II] genogroup, supported by findings of sequence conservation of binding sites and similar glycan-binding profiles between P[19] and P[10]/P[12] RVs, considered to be the earliest traceable ancestors of the P[II] genotype (17). We propose that P[6] represents an RV evolutionary intermediate in P[II] that has a preference for less mature type 1 HBGA glycans shared between animals and humans (28). Epidemiological studies showed no significant correlation between secretor status and susceptibility of P[6] RV infections (29, 30), which is supported by NMR data demonstrating that P[6] binds both LNT and LNFP I with moderate affinity and crystal structure data showing that the Secretor fucose projects away from VP8*, making minimal contacts with the binding surface (22). Human P[6] RVs have a restricted geographic prevalence and are common in African countries (31, 32), which may be due to the significantly higher rate of the Lewis-negative phenotype among African populations than in other geographic areas (3, 30), consistent with NMR data demonstrating that P[6] has extremely weak affinity to LNDFH I. In summary, the Lewis epitope may have emerged as a result of the coevolution of humans under selective pressure against P[6]/P[19] RVs and other pathogens that recognized Lewis-negative type 1 HBGAs.

P[4] and P[8] genotypes are genetically closely related and are more distantly evolved from their P[I] ancestor. P[4] and P[8] are

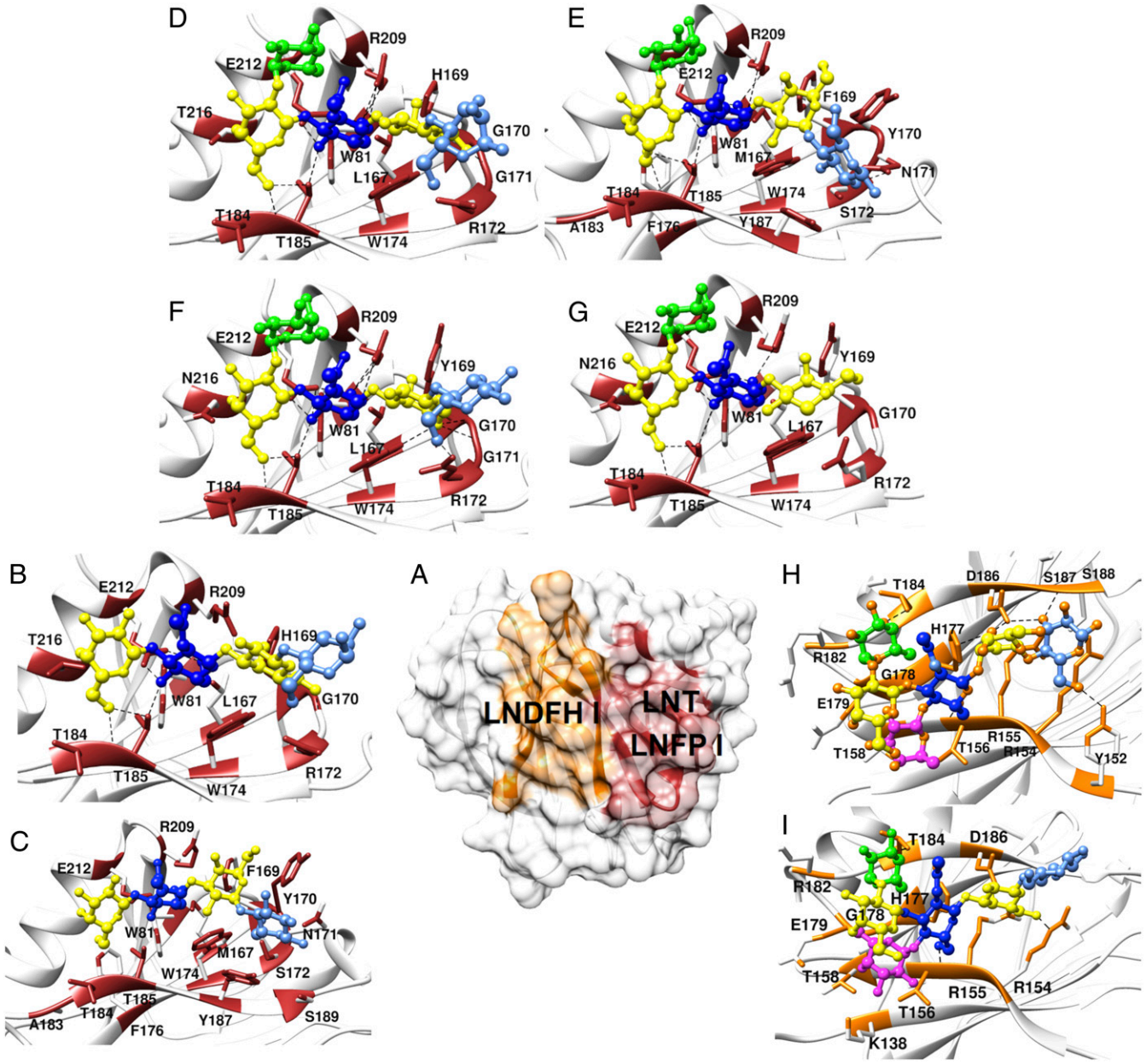

Fig. 6. Binding specificity of different P[II] VP8* genotypes for recognizing specific HBGA glycans. (*A*) Surface of the P[8] VP8* [PDB ID: 6VKX] domain showing a summary of distinct glycan-binding sites with recognition of the LNT and LNFP I using the βα site colored brown and binding to LNDFH I colored orange. (*B–I*) Ribbon diagrams highlighting binding interfaces of (*B*) P[19]-LNT [PDB ID: 5YMT], (*C*) P[6]-LNT [PDB ID: 7KI5], (*D*) P[19]-LNFP I [PDB ID: 5VKS], (*E*) P[6]-LNFP I [PDB ID: 6OAI], (*F*) P[4]-LNFP I [PDB ID: 5VX5], (*G*) P[8]-LNFP I [PDB ID:6K20], (*H*) P[8]-LNDFH I [PDB ID: 7JHZ], and (*I*) P[4]-LNDFH I [PDB ID: 7KHU]. Residues involved in the binding interface were labeled and colored brown in the βα site and orange in the ββ site. Different carbohydrate moieties of the glycan ligand were colored as follows: Lewis fucose (magenta), secretor fucose (green), galactose (yellow), *N*-acetylglucosamine (blue), and glucose (cornflower blue).

commonly found in humans but rarely in animals, consistent with the observation that P[4] and P[8] have tighter binding affinities than P[6] and P[19] to more complex HBGAs containing Lewis epitopes that are widely distributed in humans. The incorporation of Lewis epitopes into HBGAs shifts binding to the VP8* ββ site, and the Secretor fucose and Lewis fucose in the LNDFH I act as fixed points in stabilizing LNDFH I–VP8* interactions. High prevalence of these two RV genotypes may be due to the worldwide distribution of Le^b antigens. Epidemiology studies performed in multiple countries, including the United States, France, Sweden, China, Vietnam, and Burkina Faso, demonstrated that P[8] RVs infected only Secretor children (29, 30, 33–36), and biochemical and epidemiological studies demonstrated that non-Secretor individuals and individuals lacking Lewis b epitopes may be resistant to P[4] and P[8] infections (30, 37, 38). These observations agree with our NMR titration experiments showing that P[8] and P[4] did not recognize tetra-saccharide LNT without Secretor and Lewis epitopes. However, P[8] RVs were found to infect both Secretors and non-Secretors in a study on Tunisia children (39), which may be explained by the ability of P[8] to recognize type 1 precursor lacto-*N*-biose and H1 trisaccharide as shown by Gozalbo-Rovira (20).

The crystal structures showed that all human P[II] RVs employ the βα-binding site to bind the H type 1 ligand LNFP I that

Table 2. Summary of dissociation constants, K_d, obtained from NMR HSQC titration experiments

K_d (mM)	LNT	LNFP I	LNDFH I
P[19]	6.3	2.6†	48.2
P[6]	2.5*	13.6*	30.4
P[4]	no	12.5	2.5
P[8]	no	23.5*	6.3*

"No" indicates that binding was not detected for the protein–glycan pairs.
*Values were obtained from Xu et al. (22).
†Values were obtained from Liu et al. (17).

contains the Secretor fucose (18, 19, 21). Human P[19] also recognizes LNT, the apparent earliest HBGA evolutionarily ancestor lacking both the Secretor and Lewis epitopes, using the same βα-binding site (25) as does human P[8] in recognizing H type 1 trisaccharide (Fuc-α1,2-Gal-β1,3-GlcNAc) and its precursor lacto-*N*-biose (Gal-β1,3-GlcNAc) (20). Recently, P[8] was shown to bind Lewis epitope–containing HBGAs using a distinct ββ-binding site based on HSQC NMR titration experiments (22). Here, we confirmed that the addition of the Lewis epitope to the H type 1 saccharide shifts ligand binding in P[II] RVs from the βα- to the ββ-binding site. Most importantly, we obtained crystal structures of LNDFH I–bound P[4]/P[8] that provided molecular details explaining how P[4]/P[8] recognizes LNDFH I. Collectively, the emerging picture of how P[II] RV VP8* domains recognize diverse HBGA glycans derived from both our structural studies and NMR investigations provides increasing support for the hypothesis that the evolutionary adaptation of RVs was driven by selective pressure to the additional Lewis epitope.

Relative affinities of different human RV strains to different HBGAs may help understand different viral susceptibilities between Secretor and non-Secretor individuals and Lewis-positive and Lewis-negative individuals. Different affinities are caused by subtle amino acid differences within and outside the HBGA ligand–binding pockets between different RV strains as demonstrated by PCA and surface properties analyses. Previously, Böhm et al. examined the significance of HBGAs as rotavirus receptors and demonstrated the importance of A-type HBGAs for P[14] (P[III]) and P[9] (P[III]) human RV infections (40). Here, we further characterized the molecular basis for P[II] RV recognition of H type 1 and Lewis b antigens and investigated the connection between RV recognition of different HBGA phenotypes and their corresponding epidemiology. We previously demonstrated the critical role that LNDFH I plays in RV infection by showing that LNDFH I-BSA inhibits P[8] RV strain Wa infection of HT29 cells (22). Overall, the results reported here help explain the role that the synthetic stage of host HBGAs plays in defining RV host ranges, disease burden, and evolution of RV genotypes in the P[II] genogroup, which may provide valuable information for future vaccine improvements.

Materials and Methods

Expression and Purification of VP8* Proteins in *Escherichia coli*. The VP8* core fragment (amino acids 64 to 223) of human RV P[4] (strain BM5265), P[6] (strain BM11596), P[8] (strain BM13851), and P[19] (strain NIV929893) with an N-terminal glutathione S-transferase (GST) tag was overexpressed in *Escherichia coli BL21(DE3)* cells as previously described (16, 41). Luria broth medium supplemented with 100 $\mu g \cdot mL^{-1}$ ampicillin was used to grow cells at 310 K. Protein expression was induced by adding 0.5 mM isopropyl-β-D-thiogalactopyranoside to the medium when the optical density at 600 nm reached around 0.8. The cell pellet was harvested within 12 h after induction and resuspended in lysis buffer (20 mM Tris-HCl, 150 mM NaCl, and 10% glycerol pH 8.0). A French press (Thermo Fisher Scientific) was used to lyse cells, and the supernatant of the bacterial lysate was loaded to a disposable column (Qiagen) prepacked with glutathione agarose (Thermo Fisher Scientific). GST fusion protein was eluted with elution buffer (10 mM reduced glutathione and 50 mM Tris-HCl, pH 8.0), and the GST tag was removed using thrombin (Thermo Fisher Scientific) after dialysis into buffer (20 mM Tris-HCl and 50 Mm NaCl, pH 8.0). VP8* protein was further purified using size exclusion chromatography with a Superdex 200 Hiload (GE life Science) column. Purified protein was concentrated with an Amicon Ultra-10 (Millipore). ^{15}N-labeled P[4], P[6], and P[19] or ^{15}N,^{13}C-labeled P[4] was made using ^{15}N or ^{15}N,^{13}C-labeled minimal growth medium.

Crystallization, Data Collection, and Structure Determination. Hanging-drop vapor diffusion was used to crystallize human P[4] and P[8] VP8* at a concentration of 1 mM. Co-crystallization was set up with protein/ligand ratios ranging from 1:10 to 1:100 at a protein concentration of 0.2 to 0.4 mM and a ligand concentration of 4 to 40 mM. Glycans were purchased from Biosynth International, Inc. and Dextra Laboratories LTD. Crystals of P[4] VP8* were obtained by mixing 1 μL purified P[4] VP8* with 1 μL reservoir buffer: 0.1 M BIS-Tris pH 5.5 and 25% wt/vol polyethylene glycol 3350. Crystals of P[8] VP8* were obtained by mixing 1 μL purified P[8] VP8* with 1 μL reservoir buffer: 0.1 M Hepes sodium pH 7.5, 2% vol/vol polyethylene glycol 400, and 2.0 M ammonium sulfate. Crystals of LNDFH I-P[8] VP8*, LNDFH I-P[4] VP8*, and LNT-P[6] VP8* were obtained via microseed matrix screening described previously (42). Crystals of LNDFH I-P[8] and LNDFH I-P[4], obtained at 1:20 protein/ligand ratio, and LNT-P[6] at 1:40 protein/ligand ratio were harvested and immediately flash cooled in liquid nitrogen. Diffraction data were collected at the Advanced Photon Source beamline 31-ID-D, Argonne National Laboratory, in Chicago, Illinois. Images were indexed and integrated with Imosflm (43) and DIALS (44) and scaled with SCALA (45) and DIALS (46). Molecular replacement was performed with PHASER (47) using coordinates of chain A from 6NIW (22) as the search model. Restraints for LNDFH I and LNT were prepared with the small molecule topology generator PRODRG (48). Iterative model building was manually carried out in coot (49), and refinements using 5% of reflections in Free-R set were carried out in REFMAC (50) implemented in CCP4 (51). The structure quality was assessed using MolProbity (52). Final model and scaled reflection data were deposited at the Protein Data Bank (PDB): P[4] VP8* (PDB: 6UT9), P[8] VP8* (PDB: 6VKX), LNDFH I–bound P[8] VP8* (PDB: 7JHZ), LNDFH I–bound P[4] VP8* (PDB: 7KHU), and LNT-bound P[6] (PDB: 7KI5). The visualization and analysis of the final model was achieved using Chimera (53). Sequences of the VP8* domain were aligned using Clustal Omega (54).

Protein Backbone Chemical Shift Assignments. The ∼0.6 mM double-labeled ^{15}N, ^{13}C-P[4] VP8* sample was put into a 5-mm Shigemi NMR tube and spectra collected at 298 K on 600-MHz Bruker Avance III NMR spectrometer equipped with a conventional 5-mm HCN probe. Backbone chemical shifts were assigned based on the following three-dimensional spectra: HNCACB, CBCA(CO)NH, HNCO, HN(CA)CO, HNCA, and HN(CO)CA. Spectra were processed with NMRPipe (55) and analyzed with NMRFAM-SPARKY (56). Backbone ^1H, ^{13}C, and ^{15}N resonance assignments were first made using PINE (57) and then manually confirmed through NMRFAM-SPARKY. Backbone chemical shifts for P[4] were deposited into the BioMagResBank with entry identification 28109.

NMR Titration Experiments. NMR data were collected using samples in phosphate-buffered saline buffer in 5-mm Shigemi NMR tubes on 600-MHz Bruker Avance III or 850-MHz Bruker Avance II NMR spectrometers equipped with conventional 5-mm HCN probes. Chemical shift perturbations were monitored using two-dimensional ^1H-^{15}N HSQC for P[4], P[6], and P[19] VP8* upon the titration of each glycan ligand. Spectra were processed with NMRPipe (55) and analyzed with NMRFAM-SPARKY (56). Euclidean chemical shift changes upon titration were determined as the following: $\Delta\delta_{obs} = \sqrt{\frac{1}{2}[\delta_H^2 + (\alpha \cdot \delta_N^2)]}$, $\delta_H = \delta_{H\ free} - \delta_{H\ bound}$, and $\delta_N = \delta_{N\ free} - \delta_{N\ bound}$, where $\delta_{H\ free}$ and $\delta_{H\ bound}$, $\delta_{N\ free}$, and $\delta_{N\ bound}$ are backbone amide hydrogen and amide nitrogen chemical shifts in the absence and presence of ligand, respectively. The Euclidean weighting correction factor α was set to 0.14 (58). Amino acids with chemical shifts greater than threshold values determined from SDs of Euclidean chemical shift changes were used to calculate dissociation constants through global fitting. The following equation was used to extract the dissociation constant: $\Delta\delta_{obs} = \Delta\delta_{max}\left\{([P]_t + [L]_t + K_d) - [([P]_t + [L]_t + K_d)^2 - 4[P]_t[L]_t]^{\frac{1}{2}}\right\}/2P_t$, where $[P]_t$ and $[L]_t$ represent the total concentrations of protein and ligand, respectively, $\Delta\delta_{obs}$ is the chemical shift change upon titration, $\Delta\delta_{max}$ is the maximum chemical shift change on saturation, and K_d represents the dissociation constant.

PCA of the P[II] VP8*. Human P[II] VP8* sequences were obtained from the Rotavirus Classification Working Group (59) database and aligned by MUSCLE (60). The PCA of sequences was conducted with MATLAB R2019a as described previously (23). To connect sequence information contained in the first three PCA loadings with the VP8* domain structure in the P[II] genogroup, we defined categories based on the percentage of the Euler distance from the origin to the maximum Euler distance to identify amino acid positions that accounted for the most variance. Four categories of Euler distances ranging from >30% (9 residues out of 160 residues), 15 to 30% (37 out of 160 residues), 1 to 15% (39 out of 160 residues), and <1% (75 out of 160 residues) of the maximum Euler distance were assigned and mapped on the structure of VP8* domain.

Sequence Conservation Analysis of the P[II] VP8* Domain. Positional conservation of P[II] VP8* sequences was calculated as described (24) and mapped onto structures using Chimera (53). Poisson–Boltzmann electrostatic calculations that take into account spatial variations in dielectric according to molecular shape were conducted using APBS/PDB2PQR (61, 62) and mapped onto structures using Chimera (53). The superposition of LNFP I into the ββ site and LNDFH I into the βα site was performed using Chimera and gently minimized before analysis, with energy between the protein and ligand calculated using the Prodigy-Ligand server (63), the protein interface area calculated with dr-sasa (64), and hydrogen bonds identified with Chimera (53).

Data Availability. All study data are included in the article and/or *SI Appendix*.

ACKNOWLEDGMENTS. This research used resources of the Advanced Photon Source, a US Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357. Use of the Lilly Research Laboratories Collaborative Access Team beamline at Sector 31 of the Advanced Photon Source was provided by Eli Lilly Company, which operates the facility. We acknowledge the assistance of Dr. Mueller in using Miami University's High Performance Computing Center Redhawk cluster to use the DIALS (Diffraction Integration for Advanced Light Sources) software for processing crystal diffraction data. M.A.K. acknowledges support of Miami University and the Ohio Eminent Scholar program.

1. M. D. Bowen *et al*., Rotavirus strain trends during the postlicensure vaccine era: United States, 2008–2013. *J. Infect. Dis.* **214**, 732–738 (2016).
2. E. Leshem *et al*., Distribution of rotavirus strains and strain-specific effectiveness of the rotavirus vaccine after its introduction: A systematic review and meta-analysis. *Lancet Infect. Dis.* **14**, 847–856 (2014).
3. E. Heylen *et al*., Human P [6] rotaviruses from sub-Saharan Africa and Southeast Asia are closely related to those of human P [4] and P [8] rotaviruses circulating worldwide. *J. Infect. Dis.* **214**, 1039–1049 (2016).
4. G. E. Armah *et al*., Efficacy of pentavalent rotavirus vaccine against severe rotavirus gastroenteritis in infants in developing countries in sub-Saharan Africa: A randomised, double-blind, placebo-controlled trial. *Lancet* **376**, 606–614 (2010).
5. S. A. Madhi *et al*., Effect of human rotavirus vaccine on severe diarrhea in African infants. *N. Engl. J. Med.* **362**, 289–298 (2010).
6. K. Zaman *et al*., Efficacy of pentavalent rotavirus vaccine against severe rotavirus gastroenteritis in infants in developing countries in Asia: A randomised, double-blind, placebo-controlled trial. *Lancet* **376**, 615–623 (2010).
7. G. L. Plosker, Pentavalent rotavirus vaccine (RotaTeq): A review of its use in the prevention of rotavirus gastroenteritis in Europe. *Drugs* **70**, 1165–1188 (2010).
8. B. Rha *et al*., Effectiveness and impact of rotavirus vaccines in the United States – 2006-2012. *Expert Rev. Vaccines* **13**, 365–376 (2014).
9. D. Weycker, M. A. Atwood, B. Standaert, G. Krishnarajah, Public health impact of accelerated immunization against rotavirus infection among children aged less than 6 months in the United States. *Hum. Vaccin. Immunother.* **10**, 2032–2038 (2014).
10. S. Broor, D. Ghosh, P. Mathur, Molecular epidemiology of rotaviruses in India. *Indian J. Med. Res.* **118**, 59–67 (2003).
11. S. Gazal, A. K. Taku, B. Kumar, Predominance of rotavirus genotype G6P[11] in diarrhoeic lambs. *Vet. J.* **193**, 299–300 (2012).
12. Y. S. Malik *et al*., Frequency of group A rotavirus with mixed G and P genotypes in bovines: Predominance of G3 genotype and its emergence in combination with G8/G10 types. *J. Vet. Sci.* **13**, 271–278 (2012).
13. J. M. Mwenda *et al*., Burden and epidemiology of rotavirus diarrhea in selected African countries: Preliminary results from the African Rotavirus Surveillance Network. *J. Infect. Dis.* **202**, S5–S11 (2010).
14. A. Mukherjee, S. Mullick, N. Kobayashi, M. Chawla-Sarkar, The first identification of rare human group A rotavirus strain G3P[10] with severe infantile diarrhea in eastern India. *Infect. Genet. Evol.* **12**, 1933–1937 (2012).
15. P. Khamrin *et al*., Molecular characterization of VP4, VP6, VP7, NSP4, and NSP5/6 genes identifies an unusual G3P[10] human rotavirus strain. *J. Med. Virol.* **81**, 176–182 (2009).
16. Y. Liu *et al*., Rotavirus VP8*: Phylogeny, host range, and interaction with histo-blood group antigens. *J. Virol.* **86**, 9899–9910 (2012).
17. Y. Liu *et al*., Glycan specificity of P [19] rotavirus and comparison with those of related P genotypes. *J. Virol.* **90**, 9983–9996 (2016).
18. Y. Liu *et al*., Structural basis of glycan specificity of P[19] VP8*: Implications for rotavirus zoonosis and evolution. *PLoS Pathog.* **13**, e1006707 (2017).
19. L. Hu *et al*., Glycan recognition in globally dominant human rotaviruses. *Nat. Commun.* **9**, 2631 (2018).
20. R. Gozalbo-Rovira *et al*., Unraveling the role of the secretor antigen in human rotavirus attachment to histo-blood group antigens. *PLoS Pathog.* **15**, e1007865 (2019).
21. X. Sun *et al*., Structural basis of glycan recognition in globally predominant human P [8] rotavirus. *Virol. Sin.* **35**, 156–170 (2020).
22. S. Xu *et al*., Molecular basis of P[II] major human rotavirus VP8* domain recognition of histo-blood group antigens. *PLoS Pathog.* **16**, e1008386 (2020).
23. B. Wang, M. A. Kennedy, Principal components analysis of protein sequence clusters. *J. Struct. Funct. Genomics* **15**, 1–11 (2014).
24. H. Ashkenazy *et al*., ConSurf 2016: An improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res.* **44**, W344–W350 (2016).
25. X. Sun *et al*., Glycan binding specificity and mechanism of human and porcine P [6]/P [19] rotavirus VP8* s. *J. Virol.* **92**, e00538–e00518 (2018).
26. J. Le Pendu, "Histo-blood group antigen and human milk oligosaccharides" in *Protecting Infants through Human Milk*, L. K. Pickering, A. L. Morrow, G. M. Ruiz-Palacios, R. Schanler, Eds. (Springer, 2004), pp. 135–143.
27. N. Santos, Y. Hoshino, Global distribution of rotavirus serotypes/genotypes and its implication for the development and implementation of an effective rotavirus vaccine. *Rev. Med. Virol.* **15**, 29–56 (2005).
28. X. Jiang, Y. Liu, M. Tan, Histo-blood group antigens as receptors for rotavirus, new understanding on rotavirus epidemiology and vaccine strategy: Rotavirus host receptor and vaccine strategy. *Emerg. Microbes Infect.* **6**, e22 (2017).
29. A. Kambhampati, D. C. Payne, V. Costantini, B. A. Lopman, Host genetic susceptibility to enteric viruses: A systematic review and metaanalysis. *Clin. Infect. Dis.* **62**, 11–18 (2016).
30. J. Nordgren *et al*., Both Lewis and secretor status mediate susceptibility to rotavirus infections in a rotavirus genotype-dependent manner. *Clin. Infect. Dis.* **59**, 1567–1573 (2014).
31. L. M. Seheri *et al*., Characterization and molecular epidemiology of rotavirus strains recovered in Northern Pretoria, South Africa during 2003-2006. *J. Infect. Dis.* **202**, S139–S147 (2010).
32. Z. Waggie, A. Hawkridge, G. D. Hussey, Review of rotavirus studies in Africa: 1976-2006. *J. Infect. Dis.* **202**, S23–S33 (2010).
33. N. Van Trang *et al*., Association between norovirus and rotavirus infection and histo-blood group antigen types in Vietnamese children. *J. Clin. Microbiol.* **52**, 1366–1374 (2014).
34. B.-M. Imbert-Marcille *et al*., A FUT2 gene common polymorphism determines resistance to rotavirus A of the P[8] genotype. *J. Infect. Dis.* **209**, 1227–1230 (2014).
35. D. C. Payne *et al*., Epidemiologic association between FUT2 secretor status and severe rotavirus gastroenteritis in children in the United States. *JAMA Pediatr.* **169**, 1040–1045 (2015).
36. X. Ma *et al*., Binding patterns of rotavirus genotypes P[4], P[6], and P[8] in China with histo-blood group antigens. *PLoS One* **10**, e0134584 (2015).
37. B. Lee *et al*., Histo-blood group antigen phenotype determines susceptibility to genotype-specific rotavirus infections and impacts measures of rotavirus vaccine efficacy. *J. Infect. Dis.* **217**, 1399–1407 (2018).
38. X.-F. Zhang *et al*., P[8] and P[4] rotavirus infection associated with secretor phenotypes among children in South China. *Sci. Rep.* **6**, 34591 (2016).
39. S. Ayouni *et al*., Rotavirus P [8] infections in persons with secretor and nonsecretor phenotypes, Tunisia. *Emerg. Infect. Dis.* **21**, 2055–2058 (2015).
40. R. Böhm *et al*., Revisiting the role of histo-blood group antigens in rotavirus host-cell invasion. *Nat. Commun.* **6**, 5907 (2015).
41. P. Huang *et al*., Spike protein VP8* of human rotavirus recognizes histo-blood group antigens in a type-specific manner. *J. Virol.* **86**, 4833–4843 (2012).
42. A. D'Arcy, T. Bergfors, S. W. Cowan-Jacob, M. Marsh, Microseed matrix screening for optimization in protein crystallization: What have we learned? *Acta Crystallogr. F Struct. Biol. Commun.* **70**, 1117–1126 (2014).
43. H. R. Powell, The Rossmann Fourier autoindexing algorithm in MOSFLM. *Acta Crystallogr. D Biol. Crystallogr.* **55**, 1690–1695 (1999).
44. G. Winter *et al*., DIALS: Implementation and evaluation of a new integration package. *Acta Crystallogr. D Struct. Biol.* **74**, 85–97 (2018).
45. P. Evans, Scaling and assessment of data quality. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 72–82 (2006).
46. J. Beilsten-Edmands *et al*., Scaling diffraction data in the DIALS software package: Algorithms and new approaches for multi-crystal scaling. *Acta Crystallogr. D Struct. Biol.* **76**, 385–399 (2020).
47. A. J. McCoy *et al*., Phaser crystallographic software. *J. Appl. Cryst.* **40**, 658–674 (2007).
48. A. W. Schüttelkopf, D. M. van Aalten, PRODRG: A tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 1355–1363 (2004).
49. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).
50. G. N. Murshudov, A. A. Vagin, E. J. Dodson, Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* **53**, 240–255 (1997).

51. M. D. Winn *et al*., Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 235–242 (2011).
52. V. B. Chen *et al*., MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12–21 (2010).
53. E. F. Pettersen *et al*., UCSF Chimera – A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
54. F. Sievers *et al*., Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011).
55. F. Delaglio *et al*., NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277–293 (1995).
56. W. Lee, M. Tonelli, J. L. Markley, NMRFAM-SPARKY: Enhanced software for biomolecular NMR spectroscopy. *Bioinformatics* **31**, 1325–1327 (2015).
57. A. Bahrami, A. H. Assadi, J. L. Markley, H. R. Eghbalnia, Probabilistic interaction network of evidence algorithm and its application to complete labeling of peak lists from protein NMR spectroscopy. *PLoS Comput. Biol.* **5**, e1000307 (2009).
58. M. P. Williamson, Using chemical shift perturbation to characterise ligand binding. *Prog. Nucl. Magn. Reson. Spectrosc.* **73**, 1–16 (2013).
59. J. Matthijnssens *et al*., Uniformity of rotavirus strain nomenclature proposed by the Rotavirus Classification Working Group (RCWG). *Arch. Virol.* **156**, 1397–1413 (2011).
60. R. C. Edgar, MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
61. N. A. Baker, D. Sept, S. Joseph, M. J. Holst, J. A. McCammon, Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 10037–10041 (2001).
62. T. J. Dolinsky, J. E. Nielsen, J. A. McCammon, N. A. Baker, PDB2PQR: An automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res.* **32**, W665-7 (2004).
63. A. Vangone *et al*., Large-scale prediction of binding affinity in protein-small ligand complexes: The PRODIGY-LIG web server. *Bioinformatics* **35**, 1585–1587 (2019).
64. J. Ribeiro, C. Ríos-Vera, F. Melo, A. Schüller, Calculation of accurate interatomic contact surface areas for the quantitative analysis of non-bonded molecular interactions. *Bioinformatics* **35**, 3499–3501 (2019).