JOURNAL OF VIROLOGY, June 2007, p. 5949–5957 Vol. 81, No. 11
0022-538X/07/$08.00+0 doi:10.1128/JVI.00219-07

Structural Basis for the Recognition of Blood Group Trisaccharides by Norovirus▽

Sheng Cao,[1]† Zhiyong Lou,[2]† Ming Tan,[3] Yutao Chen,[1] Yijin Liu,[1] Zhushan Zhang,[1] Xuejun C. Zhang,[1,4] Xi Jiang,[3]* Xuemei Li,[1]* and Zihe Rao[1,2]

National Laboratory of Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, 15 Datun Road, Beijing 100101, China[1]; Laboratory of Structural Biology, Tsinghua University, Beijing 100084, China[2]; Division of Infectious Diseases, Cincinnati Children's Hospital Medical Center, University of Cincinnati College of Medicine, 3333 Burnet Avenue, Cincinnati, Ohio 45229-3039[3]; and Crystallography Research Program, Oklahoma Medical Research Foundation, 825 N.E. 13th Street, Oklahoma City, Oklahoma 73104[4]

Received 31 January 2007/Accepted 19 March 2007

Noroviruses are one of the major causes of nonbacterial gastroenteritis epidemics in humans. Recent studies on norovirus receptors show that different noroviruses recognize different human histo-blood group antigens (HBGAs), and eight receptor binding patterns of noroviruses have been identified. The P domain of the norovirus capsids is directly involved in this recognition. To determine the precise locations and receptor binding modes of HBGA carbohydrates on the viral capsids, a recombinant P protein of a GII-4 strain norovirus, VA387, was cocrystallized with synthetic type A or B trisaccharides. Based on complex crystal structures observed at a 2.0-Å resolution, we demonstrated that the receptor binding site lies at the outermost end of the P domain and forms an extensive hydrogen-bonding network with the saccharide ligand. The A and B trisaccharides display similar binding modes, and the common fucose ring plays a key role in this interaction. The extensive interface between the two protomers in a P dimer also plays a crucial role in the formation of the receptor binding interface.

Noroviruses, a group of small, round-structured, RNA viruses within the *Caliciviridae* family, are widespread human pathogens causing epidemic acute gastroenteritis (7). Noroviruses contain a single-stranded, positive-sense RNA genome of about 7.7 kb organized into three open reading frames (ORFs) (10). ORF1 encodes a polyprotein precursor of several nonstructural proteins. ORF2 and ORF3 encode the major capsid protein (VP1) and a minor structural protein (VP2), respectively. Recombinant VP1 self-assembles into empty virus-like particles similar to the native capsid in size and appearance (9). The function of VP2 remains unknown, although data suggest that it enhances the expression and stability of the viral capsids (1). Since noroviruses are nonenveloped, their spherical capsid surfaces contributed by VP1 contain structural determinants for host cell recognition.

The crystal structure of the prototype Norwalk virus capsid has been determined based on the evaluation of recombinant virus-like particles produced in baculovirus (19). The capsid exhibits icosahedral symmetry (T=3) and consists of 90 dimers or 180 monomers of the capsid protein (20). Each of the capsid monomers has two principle domains, the amino-terminal shell (S) domain and the carboxyl-terminal protruding (P) domain, linked by an approximately 10-residue hinge (19). The S domain possesses an eight-stranded antiparallel β sandwich structure and constitutes the interior icosahedral shell, and the P domain forms an arch-shaped protrusion emanating from the shell. The S domain is responsible for the intermolecular interaction of the viral capsid and alone can form smooth icosahedral particles, whereas the P domain is involved mainly in dimeric interactions to stabilize the capsid (2). The P domain can be further divided into two subdomains, P1 and P2 (19). P1 is more interior and exhibits a fairly conserved sequence, while P2 is located at the outer surface and exhibits a highly variable primary sequence, indicating that P2 is responsible for host interaction.

Noroviruses have been found to recognize human histo-blood group antigens (HBGAs) as receptors, and eight distinct receptor binding patterns of noroviruses have been described previously (5, 25). Human HBGAs are complex glycans present on the surfaces of red blood cells, on the epithelia of the gastrointestinal and respiratory tracts, or as free antigens in biologic fluids such as saliva, milk, and intestinal contents (15). Human HBGAs are highly polymorphic, and three major HBGA families, namely, the Lewis, secretor, and ABO families, are involved in norovirus recognition. Based on the host genetics of the three families, the eight receptor binding patterns of noroviruses can be sorted into two groups: the A/B and the Lewis (nonsecretor) binding groups (6). Direct evidence linking HBGAs to norovirus infections and illness has been obtained in volunteer studies with Norwalk virus, but such evidence for representatives of other receptor binding patterns remains lacking.

Functional analyses of norovirus capsids have been carried out in an attempt to map the receptor binding site and to

* Corresponding author. Mailing address for Xi Jiang: Division of Infectious Diseases, Cincinnati Children's Hospital Medical Center, University of Cincinnati College of Medicine, 3333 Burnet Ave., Cincinnati, OH 45229-3039. Phone: (513) 636-0119. Fax: (513) 636-7655. E-mail: jason.jiang@cchmc.org. Mailing address for Xuemei Li: National Laboratory of Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, 15 Datun Road, Beijing 100101, China. Phone: 86-10-64888556. Fax: 86-10-64872026. E-mail: lixm@sun5.ibp.ac.cn.

† These authors contributed equally to this work.

▽ Published ahead of print on 28 March 2007.

determine the receptor binding mode, mainly by site-directed mutagenesis analyses following computer modeling (24) or evolutionary trace analysis (4) based on phylogenetic relationships and the crystal structures of norovirus capsids. Nevertheless, the exact locations of receptor binding sites are still speculative, which has hindered our ability to combat this widely spreading virus.

With a view towards understanding the details of the interactions between the P domain and the receptor oligosaccharide, we have determined the crystal structures of the VA387 P domain in complexes with trisaccharides from A and B types of HBGAs. Based on the structures of the VA387 P domain-trisaccharide complexes, models for the interaction between HBGAs and other noroviruses are also proposed.

MATERIALS AND METHODS

Protein expression, purification, and crystallization. The expression construct was made by cloning the cDNA sequence encoding the P domain plus the hinge region of the VA387 norovirus into the pGEX-4T-1 vector (Amersham Bioscience), as described previously (23). Comparing the sequence of our construct with the originally submitted VP1 sequence of VA387 (GenBank accession number AAK84679) revealed a conservative, double point mutation, Thr355 to Ser and Phe375 to Leu, presumably due to the high mutation rate of the norovirus RNA genome, but the expressed protein showed no detectable difference from the wild-type one in receptor binding (23). The purification procedure for the P protein from *Escherichia coli* was described previously (23), and the purified P protein (i.e., residues 214 to 539) was concentrated to 12 mg/ml and stored in a buffer of 20 mM Tris-HCl (pH 8.0) and 20 mM NaCl in 100-μl aliquots at −80°C. The recombinant protein was crystallized by the hanging-drop vapor diffusion method at 16°C; the P protein solution was mixed with an equal volume (1.5 μl) of the precipitant solution, which contained 8% (wt/vol) polyethylene glycol (PEG) 3350 and 200 mM magnesium acetate. The best-diffracting, plate-shaped crystals grew to their maximum size of 0.1 by 0.3 by 0.6 mm in about 30 days. To determine the probable degradation of the P protein during crystallization, a native P domain crystal was washed quickly with water, dissolved in water, and analyzed by mass spectrometry (by using an AXIMA-CFR plus instrument; Shimadzu). Following the success of the crystallization of the P domain alone, the protein was cocrystallized with the blood group A trisaccharide {α-L-Fuc-(1→2)-[α-D-GalNAc-(1→3)]-D-Gal; Sigma } and the B trisaccharide {α-l-Fuc-(1→2)-[α-D-Gal-(1→3)]-D-Gal; Sigma } under crystallization conditions similar to those described above. Molar excesses of the trisaccharide between 0 and 80-fold the amount of the protein were tested, and an optimal molar excess for obtaining complex crystals was found to be about 65-fold.

Data collection and processing. For each of the native VA387 P domain and complex crystals, X-ray diffraction data were collected from a flash-cooled crystal on a Mar345 image plate (Mar-Research) after soaking the crystal in a cryoprotectant solution. The cryoprotectant solution free of trisaccharide contained 8% (wt/vol) PEG 3350, 200 mM magnesium acetate, 15% (vol/vol) glycerol, and 5% (vol/vol) PEG 400. The data were processed, scaled, and merged using the HKL2000 program package (18).

Structure determination and refinement. Phases of the P protein crystal structure were solved by the molecular replacement method by using a fragment structure (residues Phe218 to Arg530) of a single protomer from the Norwalk virus capsid (Protein Data Bank identification, 1IHM) (19) as the search model and the program Phaser (16); the molecular replacement solution had a Z score of 10.40 and an log likelihood gain of 72.12 (16). Residues that differ between the Norwalk virus capsid and the VA387 P protein were then replaced according to the VA387 amino acid sequence, and manual adjustments were carried out with the program O (11) guided by the difference electron density maps expressed by the following equations: $(2F_o - F_c)$ and $(F_o - F_c)$, where F_o is the observed structure factor and F_c is the calculated structure factor. Multiple cycles of rebuilding and refinement with the programs CNS (3) and O were then carried out. At the final stages of refinement, a composite omission map was calculated to eliminate model bias, and water molecules with good hydrogen-bonding potentials were located in peaks ($>2.5\sigma$) of the $(F_o - F_c)$ map. The upper B factor limit was set to 200 Å^2 in the refinement. Data-processing and structural refinement statistics are summarized in Table 1. The refined structures were evaluated with the program PROCHECK (13); in all three refined structures, a single residue, Asn373, was found in the disallowed region of the Ramachandran plot in the A-type complex, which appeared to be the result of direct interaction with a nearby A trisaccharide molecule. The structural analysis was performed mainly by using the program EdPDB (29). The radii of both the probe and the solvent molecule were assigned as 1.4 Å.

Protein Data Bank accession codes. The coordinates and structure factors for the P protein (2OBR), the complex with A trisaccharide (2OBS), and the complex with B trisaccharide (2OBT) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ.

RESULTS

Structural overview. The recombinant P protein (i.e., residues 214 to 539) of VA387, a Lordsdale-like GII-4 strain of norovirus, was expressed and purified from *E. coli* and crystallized. The native crystal structure was determined at a 2.2-Å resolution by using the molecular replacement method. It belongs to the $C222_1$ space group, with 1 protein molecule per asymmetric unit and a Matthews coefficient (V_M) of 2.2 Å^3/Da (~44% solvent content). Mass spectrometry analysis of the crystal sample indicated that the crystallized protein had a molecular mass of 34.4 kDa, compared with the calculated 35.9-kDa molecular mass of the recombinant protein. This observation suggested that the protein sample had been slowly digested proteolytically during the crystallization process, which might be beneficial for obtaining the high-quality crystal. Furthermore, residues Ala294 to His297, Thr371 to Asn373, and Asp391 to Asn393 in three loop regions could not be located on the electron density map when the P protein was crystallized alone, yet they became well defined when either A or B trisaccharide was added. In addition, there was no interpretable electron density for residues 531 to 535, indicating a flexible C terminus. The final model of the refined P domain monomer contained residues Thr224 to Gly530, with dimensions of 44 by 49 by 64 Å.

Similar to the Norwalk viral capsid, the P2 subdomain (residues 275 to 417) of VA387 appears to be an insertion in the P1 subdomain (residues 222 to 274 and 418 to 539) between Gly274 and Ala418 (Fig. 1A and B), which are 4.6 Å apart as measured between their Cα atoms. The P1 subdomain has two twisted β-sheets sharing strands β1 and β8 (Fig. 1B). Both β-sheets consist of purely antiparallel β strands. The smaller β-sheet has a strand order of β1-β8-β10, in which the amino terminus of β1 and the carboxyl terminus of β8 participate. The larger β-sheet has a strand order of β14-β1-β8-β13-β12-β11-β15, in which the carboxyl terminus of β1 and the amino terminus of β8 participate. There is only one well-defined α helix (α1, residues 454 to 463) in the entire P domain which is in the P1 subdomain and is involved in P domain homodimerization. The two β-sheets and the α1 helix plus the amino-terminal region (i.e., residues 227 to 237) contribute to the hydrophobic core of P1. The P2 subdomain features a six-stranded, antiparallel β-barrel of a Greek key topology with a strand order of β2-β3-β6-β5-β4-β7 and a well-packed hydrophobic core. The P1 and P2 subdomains are cushioned by a random coil-dominated region without a significant secondary structure (Fig. 1A, left).

There are 26 (8.2%) proline residues in each P monomer; 19 of them are conserved between VA387 and Norwalk virus, accounting for 16.4% of all the residues conserved between these strains. In comparison, the average frequency of proline

TABLE 1. Data collection and refinement statistics

Parameter[a]	Value for:		
	Native P protein	Complex with A trisaccharide	Complex with B trisaccharide
Data collection parameters			
Space group	$C222_1$	$C222_1$	$C222_1$
Cell dimension (Å)			
a	54.3	55.6	56.0
b	97.4	96.5	95.2
c	118.9	118.4	118.7
Resolution range (Å)	50–2.2 (2.3–2.2)	50–2.0 (2.1–2.0)	50–2.0 (2.1–2.0)
Total no. of reflections	123,065	111,148	194,305
No. of unique reflections	15,535	20,565	20,619
Completeness (%)	96.4 (83.1)	93.6 (70.3)	94.0 (85.3)
Redundancy	8.0 (4.7)	5.5 (3.0)	9.5 (5.7)
$I/\sigma(I)$	27.7 (4.0)	37.2 (2.5)	48.0 (5.8)
R_{merge} (%)	7.5 (37.6)	7.4 (56.7)	6.8 (26.3)
Refinement parameters			
No. of reflections in working set	13,890	18,870	19,203
No. of reflections in test set	722	970	978
R_{work}[b]	0.233	0.218	0.218
R_{free}[b]	0.275	0.264	0.259
No. of atoms (protein/trisaccharide/solvent)	2,314/0/226	2,385/36/209	2,385/33/236
Root mean square deviation			
Bond lengths (Å)	0.012	0.008	0.007
Bond angles (°)	1.858	1.494	1.454
Average B factor (Å^2)			
Total	41.6	52.4	36.8
Main chain	40.8	49.3	33.9
Side chain	42.3	51.5	36.4
Trisaccharide		84.8	44.1
α-Fuc		62.3	35.8
β-Gal		>85.0	41.3
α-GalNAc or α-Gal		>85.0	55.0
Solvent	40.3	60.7	46.9
Residues in the Ramachandran plot (%)			
Most favored	81.2	85.1	86.3
Allowed	17.1	14.5	13.7
Generously allowed	1.7	0.0	0.0
Disallowed	0.0	0.4	0.0

[a] *a*, *b*, and *c*, axes for cell dimensions; *I*, intensity of diffraction; R_{merge}, $\Sigma_{hkl}|I_i - I_m|/\Sigma_{hkl}I_m$, where I_i and I_m are the observed intensity and the mean intensity of related reflections, respectively. Values in parentheses correspond to the highest-resolution shell.

[b] $R_{work} = \Sigma||F_o| - |F_c||/\Sigma|F_o|$. $R_{free} = \Sigma_T||F_o| - |F_c||/\Sigma_T|F_o|$. where T is a test data set of 4.5% of total reflections randomly chosen and set aside prior to refinement.

residues in proteins from different species usually ranges between 2.1 and 6.5% (27). In addition, 36% of backbone carbonyl oxygen atoms and 24% of backbone amide groups in the final model hydrogen bond with crystallographically localized water molecules (i.e., operationally defined as those within 3.3 Å) to satisfy their hydrogen bond potential. The high proline frequency in the P domain and the extensive interactions of this domain with water molecules may partially contribute to its stability at the low secondary-structure content (3% α helices and 29% β strands).

P domain dimer. The P domain of VA387 forms a homodimer with dimensions of 57 by 63 by 69 Å. It comprises two asymmetric units related by the crystallographic two-fold *a* axis. The P dimer interface is formed mainly between the two β-barrels in the P2 subdomains (between strands β5 and β9) and between the two dyad-related regions surrounding the α1 helices in the P1 subdomains (Fig. 1A, right). The dimer interface contains a large buried surface area twisted by ~60° from the α1-α1 region to the β5-β9 contact regions, which greatly stabilizes the dimer structure.

The total area of the solvent-accessible surface (SAS) buried from each protomer during dimerization is about 1,800 Å^2. The contribution of nonpolar carbon atoms (52%) to the dimerization almost equals that of the polar atoms (48%), suggesting that both hydrophobic and polar interactions are important to the dimerization. Alternatively, there are a total of 57 residues from each protomer having SAS's buried upon dimerization; among them, 14 are charged residues (i.e., Arg, Lys, His, Asp, and Glu), 20 are noncharged polar residues (i.e., Ser, Thr, Asn, Gln, and Tyr), and the remaining 23 are hydro-

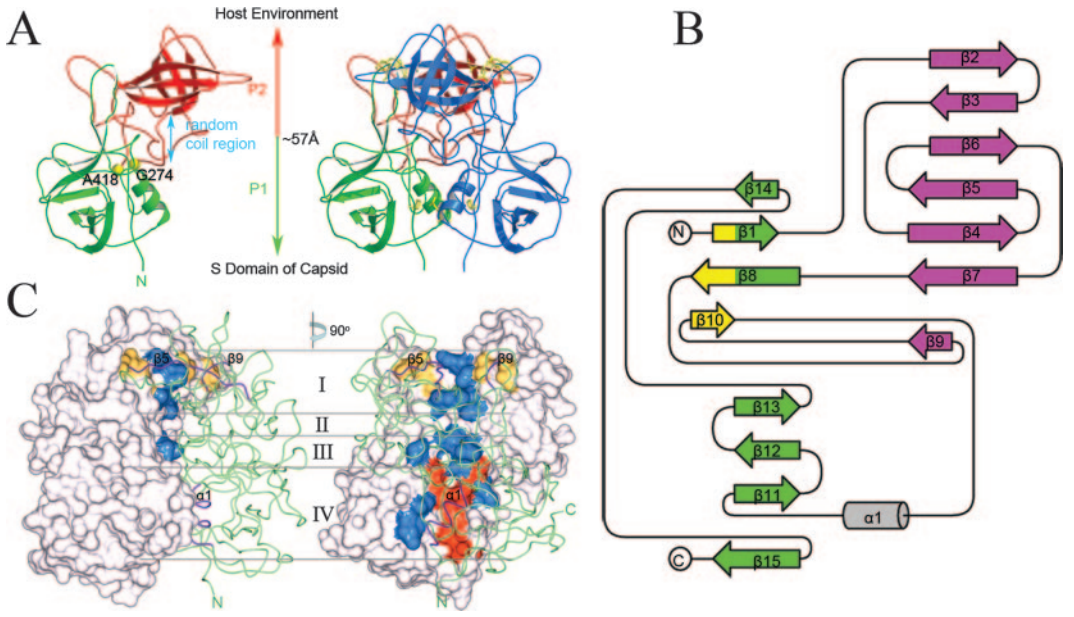

FIG. 1. Dimer of the VA387 P domain. (A) Ribbon representation of the P monomer and dimer. In the P monomer (left), the P1 subdomain is shown in green and the P2 subdomain is shown in red. The positions of Gly274 and Ala418 are marked with yellow spheres. In the P dimer (right), residue pairs (<3.3 Å apart) on the dimer interface are indicated by yellow stick models. (B) Topological diagram of the VA387 P domain. The arrows show the directions of β strands, whereas the α helix is represented by a cylinder. β strands in each β-sheet are colored identically. (C) Four layers of the dimer interface. One protomer is shown in a molecular surface model, and the other is in a light green, backbone trace model. Residues involved in hydrophobic, hydrophilic, and hydrogen-bonding interactions in the surface model are colored blue, red, and yellow, respectively. Three key secondary-structure elements on the dimer interface, β5, β9, and α1, are colored purple in the backbone model.

phobic residues. Moreover, a total of 72 water molecules (36 independent molecules corresponding to each protomer) were found in the P dimer interface (i.e., less than 6 Å from both P domain protomers and with more than 70% of the SAS buried by the P dimer). Cleary, water molecules play critical roles in stabilizing the P domain dimerization.

Along the dyad axis, we can divide the dimer interface into four layers (Fig. 1C). The top layer has a hydrophobic center and two hydrogen bond-rich regions. The hydrophobic center is formed by side chains of Met333, Lys348, Thr384, and Val386, and each hydrogen bond-rich region is formed by four intermolecular hydrogen bonds between five residues: Thr344 N-Gly442 O (2.86 Å), Thr344 Oγ-Ser441 N (2.90 Å), Thr344 Oγ-Gly442 O (2.92 Å), and Ala346 N-Cys440 O (3.17 Å). Around Arg345, there is an open cavity accessible to the receptor ligands with a size of about 10 by 8 by 4 Å (see below). The remaining three layers are a water tunnel, a hydrophobic layer (residues Pro243, Pro245, Ala280, and Val281), and a mixed hydrophilic center with two hydrophobic edges. Inside the symmetrical water tunnel in the dimer interface perpendicular to the dyad axis, with Glu274, Trp308, Thr337, and Thr436 forming the entrance, about 20 transiently located water molecules form a nearly continuous hydrogen-bonding network with the surrounding amino acid residues. The hydrophilic center of the fourth layer is formed by side chains of Ser238, Ser279, Glu455, Gln458, His459, and Glu463 as well as a number of buried water molecules, and each of the two hydrophobic edges is formed by Ile231, Leu232, and Leu278.

P domain-trisaccharide complex. The crystal forms of the P protein in complex with either the A or B trisaccharide (Fig. 2A and B) had an improved resolution (2.0 Å) and were otherwise isomorphous to that of the P domain in the absence of an oligosaccharide. The fucose (α-Fuc) and galactose (β-Gal) rings shared by the A and B trisaccharides were clearly visible at the +2.8σ contour level in $[F_{O(\text{complex})} - F_{O(\text{native})}]$ difference electron density maps phased with the protein model alone. All three saccharide rings in the trisaccharides were modeled as chair conformations. Between the two trisaccharides, the B trisaccharide was better defined in the electron density map (Fig. 2C), and the results presented below are based mainly on data from the B-type complex. It is pertinent to note, however, that the overall ligand binding modes were almost identical between the two complex crystals; there were no significant features (at ±2.8σ levels) in an $[F_{O(\text{A complex})} - F_{O(\text{B complex})}]$ difference electron density map.

Both A and B trisaccharides interacted with the P protein strongly between the fucose ring and the open cavity formed by the β5 strand, which protruded from the β-barrel, and nearby residues, including Ser343, Thr344, Arg345, and Asp374 in one protomer and Ser441, Gly442, and Tyr443 in the other. Residues Thr344, Arg345, and Gly442 were located at the bottom of the cavity, while other residues surrounded them (Fig. 3). The fact that both the A and B trisaccharides are located near the P dimer interface indicates that dimerization is not only important structurally for the P domain but also essential for its receptor binding function.

Residues Asp391 to Asn393 were missing in the native P domain crystal structure but not in those of the complexes. The binding of a trisaccharide introduced a water-bridged hydrogen bond between Asp391 and saccharides; consequently, the mobility of the Asp391-Asn393 region was reduced, and the residues became visible in the electron density map. Similarly,

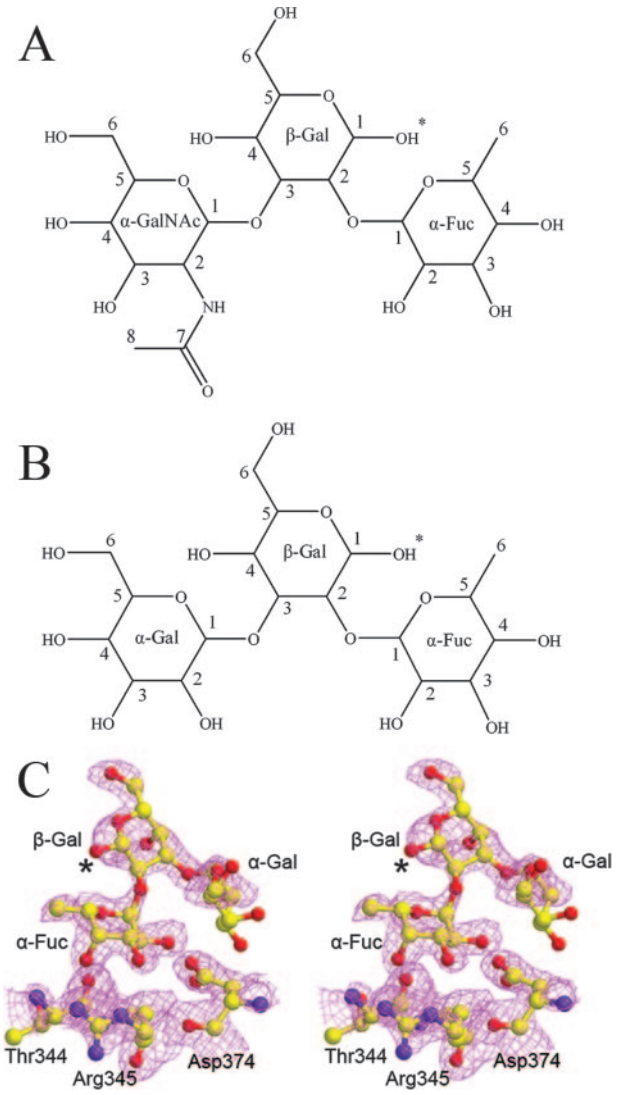

FIG. 2. Trisaccharides from the complex crystal structures. (A) Chemical structure of the A trisaccharide. The asterisk denotes the position where the trisaccharide is connected to the rest of the HBGAs. Carbon atoms circling each saccharide ring are numbered. (B) Chemical structure of the B trisaccharide. The asterisk denotes the position where the trisaccharide is connected to the rest of the HBGAs. (C) Stereo view of the B trisaccharide. The B trisaccharide and three residues from the P domain in the vicinity of the α-Fuc ring are included. The superimposed $(2F_o - F_c)$ electron density map was calculated with the trisaccharide omitted from phasing and was contoured at 0.8σ with a 2.0-Å cover radius. The asterisks correspond to that in panel B.

in the absence of a ligand, the three residues amino-terminal to Asp374 (i.e., Thr371 to Asn373) and four residues (Ala294 to His297) in a neighboring loop were missing. The binding of saccharides also requires the remodeling of the Arg345 side chain, making it suitable for interaction with the α-Fuc ring. In particular, side chains of Gln336, Thr338, and Ser343 form four hydrogen bonds with the guanidine group of Arg345 in the complex structure (Fig. 3E), while the Arg345 side chain protrudes into the solvent and is highly mobile in the native structure. A higher level of flexibility around the binding site in the absence of a ligand may be beneficial to some extent for the receptor binding by providing certain interface plasticity to accommodate different HBGAs (at a cost of entropy-related free energy of binding).

The interface between the P protein and fucose is dominated by hydrogen bonds (Fig. 3D and Table 2), and as a result the fucose ring is well defined in the electron density map. For example, both the side chain carboxyl oxygen groups of Asp374 form stable hydrogen bonds with hydroxyl groups of the fucose ring: Oδ1 to O3a (2.67 Å) and Oδ2 to O2a (2.55 Å); Asp374, in turn, is supported by His347 through a side chain hydrogen bond (2.85 Å) (Fig. 3E). The backbone O atom of Thr344 and the Nε atom of the guanidine group of Arg345 also form strong hydrogen bonds with the fucose ring: O to O4a (2.62 Å) and Nε to O3a (2.82 Å), respectively. In addition, there is a potential (3.35-Å) hydrogen bond between the Oa atom on the fucose ring and the backbone amide group of Gly442. The phenol group of the following Tyr443 exhibits a van der Waals interaction with the methyl group at position 6 of the fucose ring.

The middle β-Gal of the trisaccharide was clearly discerned in the electron density map. However, the shortest distance between the β-Gal ring and the P dimer is 6.0 Å, suggesting a certain flexibility in the mode of binding between the β-Gal ring of the trisaccharide and the P dimer. In contrast, the terminal saccharide ring that differs between A (α-GalNAc) and B (α-Gal) trisaccharides exhibited weak electron density, perhaps due to a lack of constraints from the capsid protein.

In addition to the observed ligand binding, we found another open pocket nearby the β-Gal ring. It is decorated by potential hydrogen-bonding donors and acceptors, including the backbone O and side chain Nε2 atoms of Gln390, the backbone amid group and side chain of Asp391, the backbone O atom of Gly392, and the side chains of Asn393, His395, and Tyr443. The hydroxyl group at position 1 of the β-Gal ring, which would connect to the rest of the ligand, pointed towards this open pocket. We speculate that this pocket interacts with the saccharide ring connecting to the hydroxyl group O1 of β-Gal, and such an interaction might significantly enhance the binding affinity between the saccharide and the P protein.

Comparison with Norwalk virus. The superposition of the crystal structures of the VA387 P domain and the Norwalk virus capsid yields a root mean square deviation of 1.4 Å for 245 residues from a total of 307 and 290 P domain residues for VA387 and Norwalk virus, respectively. The most significant variations in the backbones occur in the P2 subdomains, including the regions from His292 to Ile300, Thr337 to Thr344, and Thr368 to Leu375 of VA387 (Fig. 4A). Accordingly, a structure-based amino acid sequence alignment (Fig. 4B) shows an overall identity of 37%, with 46% identity in the P1 subdomain but only 24% identity in the P2 subdomain between the two viruses. In VA387, small insertions of residues Val290 to Gly295, Thr337 to Glu340, and Thr369 to Asn373 compared to the Norwalk virus sequence appear to be responsible for the different conformations in the above-mentioned regions. Notably, all three insertions are located in the turns of three β strand pairs, β2-β3, β4-β5, and β6-β7, respectively, and at the same end of the β-barrel. They form a highly variable interface between the virus and the saccharide ligands. In addition, a three-residue (Pro-Asn-Met) insertion between Try443 and Asn447 in VA387 makes the corresponding loop between β9

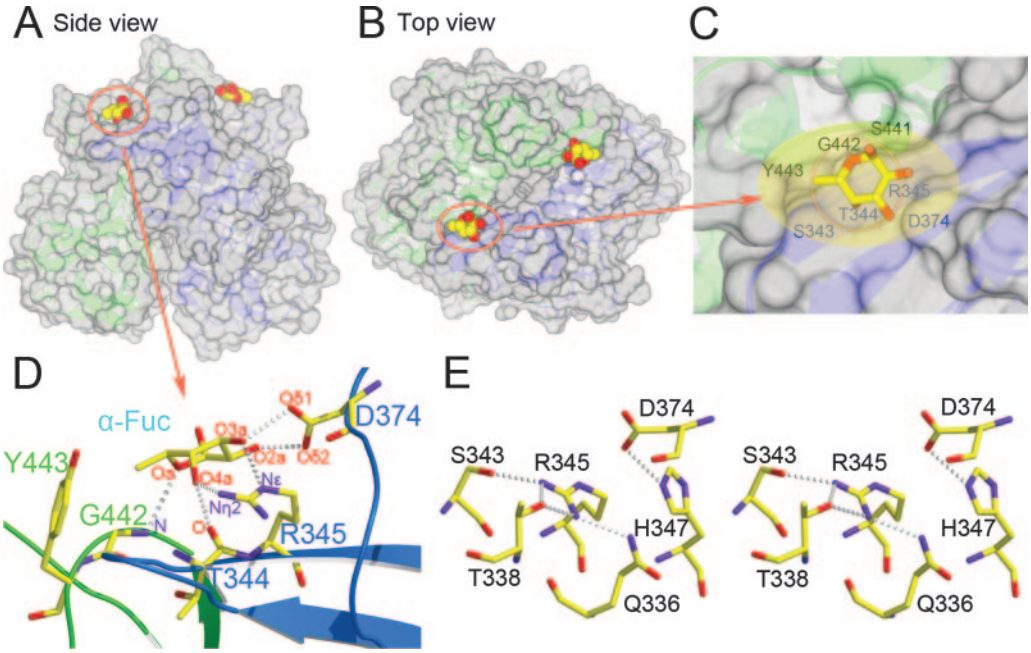

FIG. 3. α-Fuc binding site on the P domain dimer. (A and B) Side and top views of the antigen binding sites. Symmetry-related fucose rings are shown in sphere models together with the P dimer shown in the molecular surface model. (C) Surface cavity of the P dimer and the bound α-Fuc ring. The cavity interacting with the fucose ring is highlighted by the yellow ellipse, and the bottom of the cavity is circled in red. The orientation is similar to that in panel B. (D) Hydrogen-bonding network between the fucose ring and the P dimer. The fucose ring and residues involved in the interaction are shown in a ball-and-stick representation, with nitrogen, oxygen, and carbon atoms colored purple, red, and yellow, respectively. The dotted lines indicate hydrogen bonds. Backbones of the two protomers are shown as blue and green ribbons, respectively. The orientation is similar to that in panel A. Oδ1 and Oδ2, side chain groups of the Asp residue; Nϵ, Nε atom. (E) The hydrogen-bonding system for Arg345 and Asp374 on the protein side is shown in a stereo view. The orientation is similar to that in panel A.

TABLE 2. Hydrogen bond network between the B trisaccharide and the P dimer[a]

Trisaccharide oxygen group	Corresponding atom or molecule in P dimer or solvent	Distance (Å)	Bridged atom or molecule	Distance (Å)
α-Fuc				
O2a (33)	Asp374 Oδ2 (34)	2.55		
	Sol47 (41)	3.22	Ala346 N (32)	3.01
			Cys440′ O (31)	3.13
O3a (39)	Arg345 Nϵ (30)	2.82		
	Asp374 Oδ1 (37)	2.67		
O4a (35)	Thr344 O (31)	2.62		
	Arg345 Nη2 (33)	3.11		
Oa (36)	Gly442′ N (32)	3.35		
β-Gal				
O4b (47)	Sol224 (57)	2.94	Asp391′ Oδ1 (61)	2.91
O6b (30)	Sol106 (55)	2.99	Asp391′ Oδ1 (61)	3.29
α-Gal				
O2c (51)	Sol160 (61)	2.96		
	Sol109 (48)	3.22	Ser441′ Oγ (35)	2.62
			Sol47 (41)	3.03
O3c (57)	Sol185 (76)	2.49 O	Ala346 (30)	2.52
			Lys348′ Nζ (44)	2.66

[a] A 3.3-Å cutoff was used for calculating this list. Numbers in parentheses are the corresponding B factors (Å^2). Residues marked with primes are from the second P dimer protomer. In addition, there is one hydrogen bond between the trisaccharide and a nondimer neighboring P dimer molecule (Trp519 O6b-N [2.85 Å]). Solvent molecules (Sol) were numbered according to the Protein Data Bank file 2OBT.

and β10 extend into the dimer interface and results in a direct interaction of the Ser441-Tyr443 region with the saccharide ligands. Surprisingly, none of the VA387 residues interacting with the ligand trisaccharides (Table 2) are conserved in Norwalk virus, suggesting that different receptor binding modes may be used by the two viruses.

DISCUSSION

The human histo-blood group antigens have been identified as the norovirus receptors (5), and previous work on the structure of the Norwalk virus capsid has illustrated the domain organization of the VP1 protein (19). Nevertheless, conventional sequence-based homology modeling can provide only limited information on the receptor binding patterns of noroviruses because of the wide genetic diversity among norovirus sequences, which tend to undergo saturated substitution, insertions and deletions, and homologous recombination (21). Therefore, experimental, structural studies of receptor-virus complexes and subsequent site-directed mutagenesis are essential for our understanding of the specificity and mechanism of the host-virus interaction.

In the present work, we report the crystal structure of the VA387 P domain in complex with HBGA-related trisaccharides at 2.0-Å resolution, and a number of features of the P protein that are associated with viral capsid assembly and receptor binding are described. First, in the structurally observed large interface between P domain protomers, a hydrophilic cluster of charged residues (Glu455, His459, and Glu463) on

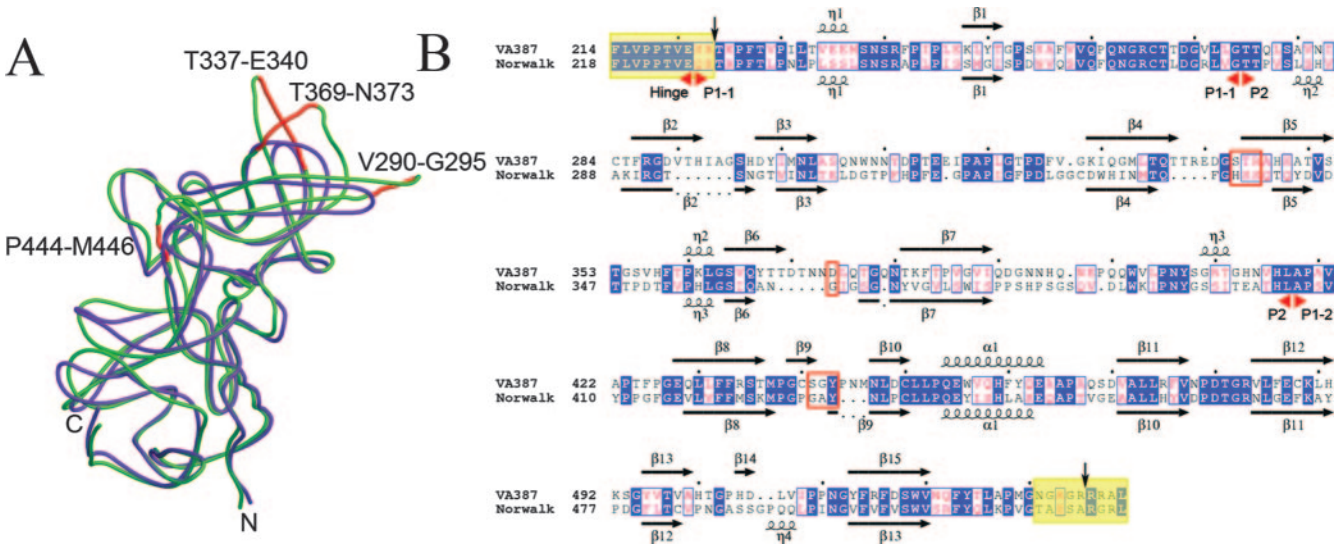

FIG. 4. Comparison of the P domains of VA387 and Norwalk virus. (A) Backbone superposition of the P protein of the VA387 strain (green) with that of Norwalk virus (purple; Protein Data Bank identification, 1IHM). The red regions represent sequence insertions (>3 amino acids) in the domain from the VA387 strain, and corresponding residue ranges are labeled along with the amino and carboxyl termini. (B) Structure-based sequence alignment of the P domain of VA387 and that of Norwalk virus. Boundaries between the hinge region, P1-1, P2, and P1-2 subdomains are marked as red triangles. Secondary-structure elements are shown for both structures, as assigned by the program DSSP (12). Identical and conserved residues are highlighted in blue and boxed with blue lines, respectively. Trisaccharide binding sites on the P protein are boxed with red lines. Regions missing in the final VA387 model are highlighted in yellow. Vertical arrows denote putative trypsin cleavage sites based on results from mass spectrometry analyses.

the α1 helix surface as well as their neighboring residues and the two dyad-symmetrical, close-contact regions between β5 and β9 form a triangular interaction network. Based on statistics on protein crystal contacts (8), this dimer interface is likely also to stabilize the homodimerization of the protein in solution, confirming the previous observations that norovirus P protein spontaneously forms stable dimers when expressed in vitro (23, 26).

Two identical, independent receptor binding sites per P dimer are identified at the outermost region of the P2 subdomain, each of which is contributed by both P protomers. The distance between the centers of the two sites is 24 Å. Among the three sugars of the trisaccharides, α-Fuc has the most extensive interaction with the P protein, consistent with the fact that this α-Fuc is shared by all secretor antigens (A, B, and H) (25). This interaction belongs to a typical mode of flat binding between saccharide and protein, as described previously (22), in which the saccharide ring lies flat on one side on the protein surface while the other side of the ring faces the solvent. Strikingly, we did not observe ligand binding in a previously predicted pocket which is at the waist of the P dimer and fairly conserved among noroviruses (24). At this time, we cannot rule out the existence of other receptor binding sites in the VA387 P dimer for the A and B or other saccharides. Such sites may have lower affinity or may have been prohibited by the crystal packing and were thus not observed in our complex crystals. Nevertheless, it is unlikely that such an additional binding site could have much higher affinity than that observed in our complex crystal; otherwise, it might have prohibited crystallization into our present crystal form.

The characteristic terminal groups for A (α-GalNAc) and B (α-Gal) trisaccharides (Fig. 2) contribute little to the binding with the P dimer in their corresponding complex crystal structures. This observation is consistent with results from a previous in vitro study, using saliva samples and synthetic oligosaccharides, showing that A and B antigens have similar binding affinities to the VA387 norovirus (23). In addition, we were unable to cocrystallize the P domain with H trisaccharide [α-Fuc-(1→2)-β-Gal-(1→4)-GlcNAc; Sigma] (data not shown). In contrast to A and B trisaccharides, the H trisaccharide possesses a GlcNAc ring at position 1 of β-Gal but contains no saccharide connected to position 3 of β-Gal. Crystal-packing analysis suggests that the GlcNAc unit should not prevent cocrystallization. Therefore, it appears that the absence of α-GalNAc or α-Gal in the H trisaccharide somehow reduces the affinity for the VA387 P domain, consistent with the in vitro binding results obtained using saliva (23). The findings of previous nuclear magnetic resonance and crystallographic studies suggest that the conformation of the B trisaccharide is more rigid in general (17). Interestingly, in our case, the conformations of both α-Fuc and β-Gal of the B trisaccharide were well defined in the electron density map and were very similar to their counterparts in the H trisaccharide in a previously described complex with *N*-acetylgalactosaminyl transferase (14).

Based the results of previous and present studies, the recognition of norovirus by HBGAs may occur at two levels. The first level is strain specific, in which a hydrogen-bonding network is formed between saccharides and residues equivalent to Thr344, Arg345, Asp374, and Gly442 of VA387. It has been shown previously that a Thr338-to-Ala mutation in VA387 abolishes the binding with HBGAs of all A, B, and O types (24). Consistently, our complex structure illustrates that Thr338 forms two hydrogen bonds with the guanidine group of

Arg345 (Fig. 3E), essential for the latter to assume an active conformation for receptor binding.

The second level of recognition is nonspecific, which involves mainly long-distance interaction necessary for initiating or stabilizing the interaction. In our cocrystallization protocol, a molar excess of trisaccharide of more than 60-fold the amount of protein was used in order to obtain high-quality complex crystals, indicating a low affinity between the P protein and the trisaccharide. In contrast, the natural HBGAs have high affinities for the viruses, indicating the existence of additional interactions beyond those involving trisaccharide. As an almost universal feature associated with viruses (28), both the native HBGAs on the cell surface and the viral capsid are multivalent, which may result in a super network at both levels of interactions and a high affinity between the host cell and the virus. In addition, parts of the glycan chains other than the terminal region of the HBGA may stabilize the overall interaction through multiple contacts. It has been noted previously that synthetic oligosaccharides conjugated to different backbones (e.g., bovine serum albumin [23] and polyphenylacetylene) reveal significantly higher affinities than free saccharides (M. Tan and X. Jiang, unpublished data), suggesting that the backbones of HBGAs may similarly be involved in the second level of interaction. The biological significance of the second-level interaction remains to be firmly established.

Norwalk virus and VA387 belong to different genogroups, and they are the only two noroviruses with their P domain three-dimensional structures resolved. While they show overall structural similarity, the compositions and structures of the receptor binding interfaces of the two viruses are significantly different. For instance, a three-residue deletion between β9 and β10 in the Norwalk virus results in the loss of one side wall of the VA387-like open pocket. This variation may shift the binding site away from the dimer interface, although the dimerization may still support the receptor binding site. A shifted but similar distribution of polar residues and backbone groups on the top surface of the Norwalk virus P dimer has been identified as a potential α-Fuc binding site. In addition, like Arg345 in VA387, Norwalk virus Gln334 may form the bottom of the putative binding pocket and make crucial hydrogen bonds with the saccharide.

In addition, VA207 shows the closest genetic relationship to VA387 in the P domain among strains that recognize nonsecretors as well as A, B, and O secretors. Primary sequence alignment shows that although the residue equivalent to VA387 Asp374 is missing in VA207, most other residues involved in the hydrogen-bonding network with α-Fuc are conserved. The function of Asp374 may be performed by a neighboring asparagine residue in VA207. A mutagenesis analysis would be necessary to verify the role of such VA207 residues in ligand binding.

Furthermore, the crystal structure of VA387 sheds light on the structural bases of the receptor binding profiles of other noroviruses, such as Snow Mountain virus, norovirus strain MOH, and Hawaii virus (Fig. 5). These strains represent three receptor binding patterns distinct from that of VA387 but share the ability to bind to the B oligosaccharide chain and do not bind to nonsecretors (Le^a, an antigen without α-Fuc linked to β-Gal) (6). Their residues corresponding to Arg345 and Asp374 of VA387 are conserved and form direct hydrogen bonds with α-Fuc. Moreover, the residues Gln336 and His347, which stabilize Arg345 and Asp374 in VA387, respectively, are also conserved. All of these points suggest that the three strains have receptor binding modes similar to that of VA387 at the α-Fuc binding site.

	336	345	374	442
SMV	... T Q ...	A N R G H ...	D D ...	G G Y ...
MOH	... S Q ...	A N R A H ...	N D ...	G G F ...
Hawaii	... S Q ...	T C R A H ...	S D ...	G G T ...
VA387	... T Q ...	S T R A H ...	N D ...	S G Y ...
VA207	... S Q ...	A T R A H ...	N - ...	S G H ...

FIG. 5. Alignment of sequences from strains representative of five binding patterns. Numbers denote the residue location in the primary sequence of VA387, where the key residues form the open binding cavity for α-Fuc. GenBank accession numbers are as follows: Snow Mountain virus (SMV), AAB61685; norovirus strain MOH, AAK84404; Hawaii virus, AAB97768; VA387, AAK84679; and VA207, AAK84676.

A full understanding of the antigen-virus recognition system would provide a fundamental road map for the future development of novel pharmaceutical interventions to prevent and treat norovirus-associated diseases. In particular, structural analyses of P domain-trisaccharide complexes will be helpful in the design of small-molecule reagents that exclusively interact with norovirus. Furthermore, such studies would stimulate research on other HBGA-related pathogens.

ACKNOWLEDGMENTS

We thank Zhijie Liu, Neil Shaw, and Mark Bartlam for technical instructions and helpful discussions. We are also very grateful to staff members of the Structural Biology Core Facility in the Institute of Biophysics, Chinese Academy of Sciences, for their excellent technical assistance, especially to Zhongnian Zhou, Meirong Zhang, Yi Han, and Xiaoxia Yu.

This work was supported by "863" Project (China) grant 2006AA02A322 to X.L., CAS (China) grant KSCX2-YW-05 to Z.R., and National Institutes of Health (United States) grant AI055649 to X.J.

REFERENCES

1. **Bertolotti-Ciarlet, A., S. E. Crawford, A. M. Hutson, and M. K. Estes.** 2003. The 3′ end of Norwalk virus mRNA contains determinants that regulate the expression and stability of the viral capsid protein VP1: a novel function for the VP2 protein. J. Virol. **77:**11603–11615.
2. **Bertolotti-Ciarlet, A., L. J. White, R. Chen, B. V. Prasad, and M. K. Estes.** 2002. Structural requirements for the assembly of Norwalk virus-like particles. J. Virol. **76:**4044–4055.
3. **Brunger, A. T., P. D. Adams, G. M. Clore, W. L. DeLano, P. Gros, R. W. Grosse-Kunstleve, J. S. Jiang, J. Kuszewski, M. Nilges, N. S. Pannu, R. J. Read, L. M. Rice, T. Simonson, and G. L. Warren.** 1998. Crystallography and NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr. D **54:**905–921.
4. **Chakravarty, S., A. M. Hutson, M. K. Estes, and B. V. Prasad.** 2005. Evolutionary trace residues in noroviruses: importance in receptor binding, antigenicity, virion assembly, and strain diversity. J. Virol. **79:**554–568.
5. **Huang, P., T. Farkas, S. Marionneau, W. Zhong, N. Ruvoen-Clouet, A. L. Morrow, M. Altaye, L. K. Pickering, D. S. Newburg, J. LePendu, and X. Jiang.** 2003. Noroviruses bind to human ABO, Lewis, and secretor histo-blood group antigens: identification of 4 distinct strain-specific patterns. J. Infect. Dis. **188:**19–31.
6. **Huang, P., T. Farkas, W. Zhong, M. Tan, S. Thornton, A. L. Morrow, and X. Jiang.** 2005. Norovirus and histo-blood group antigens: demonstration of a wide spectrum of strain specificities and classification of two major binding groups among multiple binding patterns. J. Virol. **79:**6714–6722.
7. **Hutson, A. M., R. L. Atmar, and M. K. Estes.** 2004. Norovirus disease: changing epidemiology and host susceptibility factors. Trends Microbiol. **12:**279–287.
8. **Janin, J., and F. Rodier.** 1995. Protein-protein interaction at crystal contacts. Proteins **23:**580–587.
9. **Jiang, X., M. Wang, D. Y. Graham, and M. K. Estes.** 1992. Expression,

self-assembly, and antigenicity of the Norwalk virus capsid protein. J. Virol. **66:**6527–6532.
10. **Jiang, X., M. Wang, K. Wang, and M. K. Estes.** 1993. Sequence and genomic organization of Norwalk virus. Virology **195:**51–61.
11. **Jones, T. A., J. Y. Zou, S. W. Cowan, and M. Kjeldgaard.** 1991. Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Crystallogr. A **47:**110–119.
12. **Kabsch, W., and C. Sander.** 1983. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers **22:**2577–2637.
13. **Laskowski, R. A., M. W. MacArthur, D. S. Moss, and J. M. Thornton.** 1993. PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Crystallogr. **26:**283–291.
14. **Letts, J. A., N. L. Rose, Y. R. Fang, C. H. Barry, S. N. Borisova, N. O. Seto, M. M. Palcic, and S. V. Evans.** 2006. Differential recognition of the type I and II H antigen acceptors by the human ABO(H) blood group A and B glycosyltransferases. J. Biol. Chem. **281:**3625–3632.
15. **Marionneau, S., A. Cailleau-Thomas, J. Rocher, B. Le Moullac-Vaidye, N. Ruvoen, M. Clement, and J. Le Pendu.** 2001. ABH and Lewis histo-blood group antigens, a model for the meaning of oligosaccharide diversity in the face of a changing world. Biochimie **83:**565–573.
16. **McCoy, A. J., R. W. Grosse-Kunstleve, L. C. Storoni, and R. J. Read.** 2005. Likelihood-enhanced fast translation functions. Acta Crystallogr. D **61:**458–464.
17. **Otter, A., R. U. Lemieux, R. G. Ball, A. P. Venot, O. Hindsgaul, and D. R. Bundle.** 1999. Crystal state and solution conformation of the B blood group trisaccharide alpha-L-Fucp-(1→2)-[alpha-D-Galp-(1→3)]-beta-D-Galp-OCH3. Eur. J. Biochem. **259:**295–303.
18. **Otwinowski, Z., and W. Minor.** 1997. Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. **276:**307–326.
19. **Prasad, B. V., M. E. Hardy, T. Dokland, J. Bella, M. G. Rossmann, and M. K. Estes.** 1999. X-ray crystallographic structure of the Norwalk virus capsid. Science **286:**287–290.
20. **Prasad, B. V., R. Rothnagel, X. Jiang, and M. K. Estes.** 1994. Three-dimensional structure of baculovirus-expressed Norwalk virus capsids. J. Virol. **68:**5117–5125.
21. **Rohayem, J., J. Munch, and A. Rethwilm.** 2005. Evidence of recombination in the norovirus capsid gene. J. Virol. **79:**4977–4990.
22. **Stehle, T., and S. C. Harrison.** 1997. High-resolution structure of a polyomavirus VP1-oligosaccharide complex: implications for assembly and receptor binding. EMBO J. **16:**5139–5148.
23. **Tan, M., R. S. Hegde, and X. Jiang.** 2004. The P domain of norovirus capsid protein forms dimer and binds to histo-blood group antigen receptors. J. Virol. **78:**6233–6242.
24. **Tan, M., P. Huang, J. Meller, W. Zhong, T. Farkas, and X. Jiang.** 2003. Mutations within the P2 domain of norovirus capsid affect binding to human histo-blood group antigens: evidence for a binding pocket. J. Virol. **77:**12562–12571.
25. **Tan, M., and X. Jiang.** 2005. Norovirus and its histo-blood group antigen receptors: an answer to a historical puzzle. Trends Microbiol. **13:**285–293.
26. **Tan, M., and X. Jiang.** 2005. The p domain of norovirus capsid protein forms a subviral particle that binds to histo-blood group antigen receptors. J. Virol. **79:**14017–14030.
27. **Tekaia, F., E. Yeramian, and B. Dujon.** 2002. Amino acid composition of genomes, lifestyles of organisms, and evolutionary trends: a global picture with correspondence analysis. Gene **297:**51–60.
28. **Wickham, T. J., R. R. Granados, H. A. Wood, D. A. Hammer, and M. L. Shuler.** 1990. General analysis of receptor-mediated viral attachment to cell surfaces. Biophys. J. **58:**1501–1516.
29. **Zhang, X.-J., and B. W. Matthews.** 1995. EDPDB: a multifunctional tool for protein structure analysis. J. Appl. Crystallogr. **28:**624–630.