High-resolution cryo-EM structures of outbreak strain human norovirus shells reveal size variations

James Jung[a,b], Timothy Grant[c], Dennis R. Thomas[a], Chris W. Diehnelt[d], Nikolaus Grigorieff[c,e,f], and Leemor Joshua-Tor[a,b,1]

[a]W. M. Keck Structural Biology Laboratory, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724; [b]Howard Hughes Medical Institute, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724; [c]Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, VA 20147; [d]Biodesign Institute Center for Innovations in Medicine, Arizona State University, Tempe, AZ 85281; [e]RNA Therapeutics Institute, University of Massachusetts Medical School, Worcester, MA 01605; and [f]Howard Hughes Medical Institute, University of Massachusetts Medical School, Worcester, MA 01605

Noroviruses are a leading cause of foodborne illnesses worldwide. Although GII.4 strains have been responsible for most norovirus outbreaks, the assembled virus shell structures have been available in detail for only a single strain (GI.1). We present high-resolution (2.6- to 4.1-Å) cryoelectron microscopy (cryo-EM) structures of GII.4, GII.2, GI.7, and GI.1 human norovirus outbreak strain virus-like particles (VLPs). Although norovirus VLPs have been thought to exist in a single-sized assembly, our structures reveal polymorphism between and within genogroups, with small, medium, and large particle sizes observed. Using asymmetric reconstruction, we were able to resolve a Zn^{2+} metal ion adjacent to the coreceptor binding site, which affected the structural stability of the shell. Our structures serve as valuable templates for facilitating vaccine formulations.

norovirus | cryo-EM | foodborne illnesses

Noroviruses are a leading cause of foodborne illnesses, accounting for 58% of all outbreaks and over 96% of non-bacterial outbreaks, causing ~21 million cases in the United States and 685 million cases worldwide each year (1–3). Also commonly referred to as stomach flu or winter vomiting bug, the viruses cause frequent outbreaks of acute gastroenteritis in hospitals, nursing homes, day care centers, schools, and restaurants, and over 90% of diarrheal outbreaks on cruise ships. Norovirus illnesses are estimated at a cost of $2 billion for treatment and lost productivity in the United States and $60 billion worldwide annually (1–3). The viruses are highly contagious, with as few as 18 virus particles needed to cause infection (4), and spread by the fecal–oral route through direct contact with patients and aerosolized viruses from vomiting and contaminated surfaces, foods, and water supplies (5, 6).

Noroviruses are round, nonenveloped viruses with (+)ssRNA genomes that belong to the Caliciviridae family and are divided into at least 6 genogroups that are subdivided into 30 or more genotypes, of which genogroups I, II, and IV cause illnesses in humans (5, 6). The norovirus genomes encode two structural proteins, one major capsid protein (VP1) that forms the icosahedral shell enclosing the genome and a minor structural protein (VP2) that is positively charged and may interact with and stabilize its genome (7, 8). More recently, the VP2 of feline vesivirus was shown to form a portal structure upon binding of the feline receptor (9).

The major capsid protein consists of a shell (S) domain that forms the icosahedral enclosure and a protruding (P) domain that forms dimeric spikes on the virus surface and bears the antigenic features involved in host interactions (Fig. 1). The P domain consists of two subdomains: P1 emerging from the S domain and P2, which is an insertion within P1 and positioned at the outermost surface of the virus (*SI Appendix,* Fig. S1). The P2 subdomains have been shown to bind human histo-blood group antigens (HBGAs) as attachment factors, although some strains

including GII.2 Snow Mountain virus (SMV) show only weak or no binding to HBGAs (5, 10, 11).

Recombinant major capsid proteins assemble into virus-like particles (VLPs) that preserve the structure and antigenicity of infectious virions (7, 12). VLPs and animal noroviruses have been used as model systems, due to difficulties in propagating human virions in cell-culture systems (7, 13). There are no approved treatments available for norovirus infections, and VLPs are currently being used as candidates for vaccine trials; one candidate has reached a phase II clinical trial (12). Although crystal structures of P domains are available (10, 14), currently there is only one high-resolution VLP crystal structure of one norovirus strain (GI.1) available, which was solved at 3.4-Å resolution (7). The GI.1 VLP exhibits T = 3 icosahedral symmetry, an assembly of 180 subunits, with 90 dimeric spikes on its surface.

To further our understanding of human norovirus capsid architectures and the relationship to disease, we report high-resolution

Significance

Despite being a leading cause of foodborne illnesses, accounting for 58% of all outbreaks and over 96% of nonbacterial outbreaks, there are no approved treatments available for norovirus infections. Assembled shells of the viruses without genetic materials enclosed are currently being used as candidates for vaccine trials. Although the virus shells have been thought to exist in a single-sized assembly, our structures in near-atomic detail reveal clear variations in size between different outbreak strains, and in spatial and angular arrangements of the antigenic surface spikes. The structures we present serve as valuable templates for facilitating vaccine formulations.
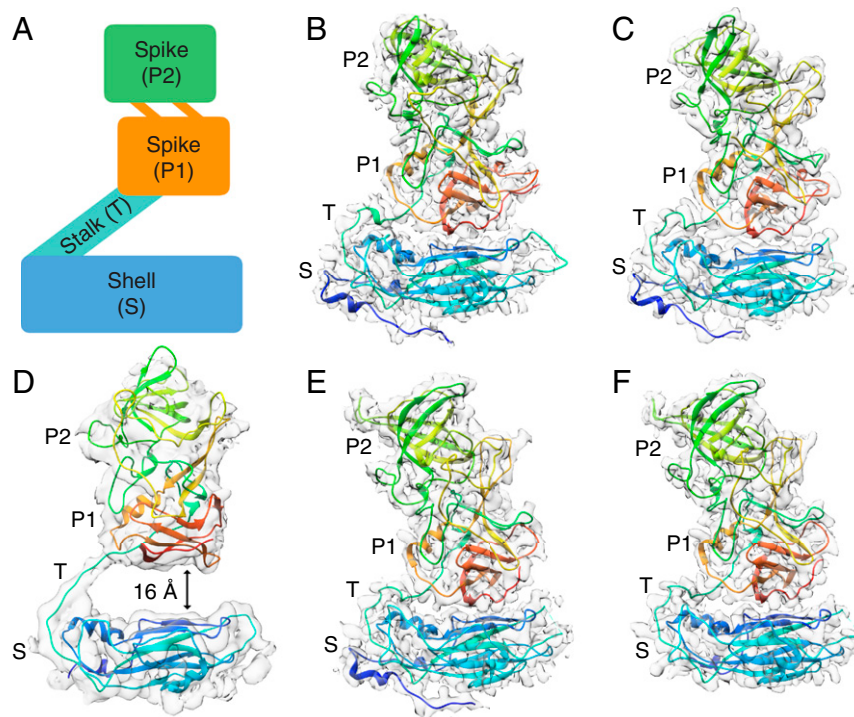
BIOPHYSICS AND COMPUTATIONAL BIOLOGY

Fig. 1. Modular organization of norovirus capsid proteins. Side views of human norovirus major capsid protein A subunits, showing the cryo-EM maps in gray and the fitted atomic models colored in rainbow representation from N to C termini. (*A*) A schematic diagram showing the modular organization of capsid subunits, consisting of trapezoid-shaped shell, long and flexible stalk, and protruding spike domains. The P domain consists of a P1 subdomain emerging from the S domain and P2, which is an insertion in P1 and positioned at the outermost surface of the virus. (*B*) GI.1 subunit, showing the P domain placed immediately above and forming contacts with the S domain. (*C*) GI.7 subunit, with the P domain placed close to the S domain. (*D*) GII.4 subunit, with the P domain lifted significantly (~16 Å) above the shell domain through the long and flexible stalk region. (*E*) GII.2 T = 3 particle subunit, with the P domain making contacts with the S domain. (*F*) GII.2 T = 1 particle subunit, also with the P domain placed close to the S domain.

cryo-EM VLP structures of four outbreak strains in the frozen hydrated state: GII.2 SMV (2.7 and 3.1 Å), GII.4 Minerva (4.1 Å), GI.7 Houston (2.9 Å), and GI.1 Norwalk (2.6 Å) strains (Fig. 2 and *SI Appendix*, Fig. S2 and Table S1). Notably, GII.4 strains have been responsible for 70–80% of all norovirus outbreaks, representing the most epidemiologically prevalent strain (5, 6). The VLP structures reveal polymorphism with T = 1, T = 3, and T = 4 assemblies observed. These structures serve as valuable templates for facilitating VLP vaccine formulations.

Results and Discussion

Shell Domain and Size Variation. Movies of VLPs embedded in vitreous ice on holey carbon grids were collected using a Titan Krios G3 transmission electron microscope (ThermoFisher) equipped with a K2 Summit direct electron detector (Gatan) (*Materials and Methods* and *SI Appendix*, Fig. S3). To improve the maps, we used an asymmetric focused reconstruction method that resolved features lost with symmetry imposition, using the program symmetry_expand_stack_and_par included with *cis*TEM).

Although norovirus VLPs have been thought to exist only in T = 3 assemblies, our cryo-EM VLP structures revealed structural polymorphism between and within genogroups in identical buffer conditions (15, 16). Our GI.1 structure is in good agreement with the previous GI.1 crystal structure (0.8-Å rmsd). The GI.7 and GI.1 strains are also both in T = 3 assemblies (Fig. 2 and *SI Appendix*, Fig. S2). The external diameters of GI.7 and GI.1 particles are 420 and 410 Å, respectively, and the internal diameter of both strains is 240 Å. The GII.4 VLPs, on the other hand, are exclusively in T = 4 icosahedral symmetry consisting of 120 spikes and 240 subunits. Accordingly, the larger particles are 490 Å in external diameter and 280 Å in internal diameter. The

GII.2 strain particles are observed in a T = 3 form as well as in T = 1 symmetry, the latter of which is likely to exist as an empty subviral form, with 30 spikes and 60 subunits. The external diameters of GII.2 T = 3 and T = 1 particles are 430 and 310 Å, respectively, and the internal diameters are 240 and 120 Å, respectively. We note that the T = 3 and T = 1 structures were observed together and solved from micrographs of the same sample.

The observed variances in the T numbers of norovirus VLPs are examples of quasi-equivalence, with identical capsid subunits undergoing conformational adjustments to fit multiple symmetry arrangements (17). Each T = 1 particle icosahedral asymmetric unit (ASU) consists of one subunit (A), whereas the T = 3 particle ASU consists of A, B, and C subunits in a quasi-threefold symmetry arrangement (Fig. 3). The T = 4 particle ASU consists of four subunits, A, B, C, and D, with the C–D dimer taking up a position similar to the C–C dimer of T = 3 particles. In the T = 4 particles, the icosahedral threefold axes are formed between three D subunits (Fig. 2 and *SI Appendix*, Fig. S2). In the T = 3 particles, B and C subunits interdigitate around the icosahedral threefold axes in quasi-sixfold symmetry. The T = 4 particle icosahedral twofold symmetry axes are formed between two sets of B, C, and D subunits in quasi-sixfold arrangements. The C–C dimer interfaces form the icosahedral twofold axes in T = 3 particles.

In the T = 3 particles of GII.2, GI.7, and GI.1, the two S domains of A–B dimers are in a bent arrangement and the C–C dimers are in plane (*SI Appendix*, Fig. S4). The N-terminal arm of GI.1 VP1 on the inner surface of the S domain has been suggested as a potential molecular switch that transitions between ordered and disordered states, where the ordered arm from the bent A–B dimer interacts with and stabilizes the neighboring flat C–C dimer conformation during T = 3 particle
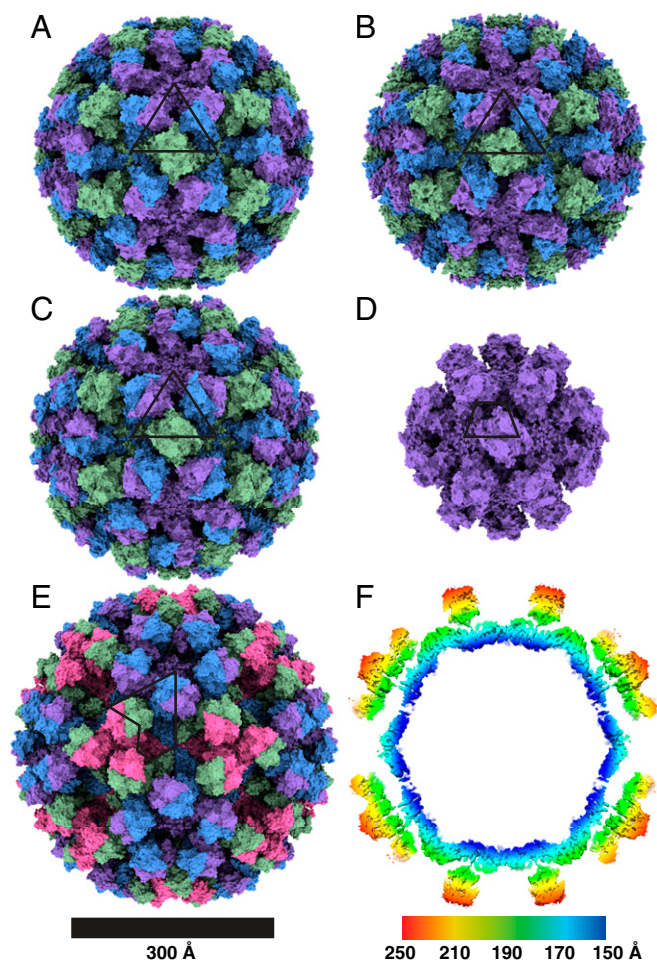
Fig. 2. Cryo-EM structures of human norovirus outbreak strain capsids. Shaded depth-cue representations of norovirus VLP structures viewed along the icosahedral twofold axis, colored purple (subunit A), blue (subunit B), green (subunit C), and pink (subunit D). The positions of asymmetric units are identified by black lines. (*A*) GI.1 Norwalk strain at 2.9-Å resolution in T = 3 icosahedral symmetry with 90 dimeric P-domain spikes assembled from 180 subunits and 410 Å in diameter. (*B*) GI.7 Houston strain at 2.9-Å resolution in T = 3 symmetry and 420 Å in diameter. (*C*) GII.2 Snow Mountain virus strain at 3.1-Å resolution in T = 3 symmetry and 430 Å in diameter. The spikes of GII.2 SMV are twisted ~15° counterclockwise relative to GI.1 and GI.7. (*D*) GII.2 SMV at 2.7-Å resolution in T = 1 symmetry with 30 spikes and 60 subunits and 310 Å in diameter. The spikes are placed significantly farther apart (~10–25 Å) in T = 1 symmetry. (*E*) GII.4 Minerva strain at 4.1-Å resolution in T = 4 symmetry with 120 spikes and 240 subunits and 490 Å in diameter. The spikes of GII.4 are twisted ~50° clockwise relative to the orientation of GI.1 and GI.7. (*F*) A central slice view of a GII.4 cryo-EM map colored by radial distance from the center (blue to red), showing a two-layered architecture with a secondary layer of spikes suspended ~16 Å above the primary layer of the icosahedral shell domain.

assembly (7, 17, 18). In the morphologically closely related tomato bushy stunt virus capsid with T = 3 symmetry, the arm of the C subunit wedges the two S domains apart in the C–C dimer interface to control the bent-to-flat conformation switching (17, 19). In our cryo-EM structures of GII.2, GI.7, and GI.1, the arm is ordered only on the B subunit that latches onto the neighboring C subunit.

In the large, T = 4, GII.4 particles, the two S domains of A–B dimers are bent as in the A–B dimers of T = 3 particles, and the C–D dimers are flat like the C–C dimers of those particles (*SI Appendix*, Fig. S4). However, the arm is disordered in all four subunits. It appears, therefore, that an ordered arm is not required

to stabilize the flat dimer conformation in large norovirus particles. Indeed, truncation of the arm still resulted in self-assembly of GI.1 T = 3 particles, and the arm was dispensable for stabilization of the flat dimers in a previous study (18). In GII.2 small, T = 1, particles, all VP1 dimers are bent and the arms are disordered.

Protruding Spike Domain. Variations in the spatial and angular orientations of subunits and their domains would have consequences in host interactions, such as bivalent Ig binding. The P domain of GII.4 is lifted significantly (~4–28 Å) from the S domain through a long and flexible stalk region (Fig. 1). The P-domain densities were not as well resolved as those of the S domain. This may be due to the greater flexibility between the P domains and the shell in this assembly. Interactions between the elevated P domains form a double-layered structure with a distinct layer of spikes suspended above the icosahedral S-domain layer (Fig. 2). The dimeric P-domain spikes of A–B and C–D dimers are tilted toward one side (*SI Appendix*, Fig. S4). The distance between P and S domains is 16 and 24 Å on the A and B subunits, respectively, and 4 and 28 Å on the C and D subunits, respectively. On the other hand, the P domains of GII.2, GI.7, and GI.1 are placed close to the S domains, and the P1 subdomain makes contacts with the S domains of its own and of a neighboring subunit (Fig. 1). The lifted P-domain positioning and secondary-layer formation do not appear to be universally shared features among genogroup II strains. The P-domain spikes of GII.2 are twisted ~25° counterclockwise relative to the orientation of GI.1 and GI.7 (Fig. 2). The spikes of GII.4 are twisted ~40° clockwise relative to GI.1 and GI.7. Similarly lifted, twisted, and tilted P-domain positioning and orientations were previously reported in an 8-Å cryo-EM reconstruction of murine norovirus 1 native infectious particles in T = 3 form (13, 20).

Recently reported cocrystal structures of murine norovirus P domains in complex with their protein receptor, CD300lf, revealed the receptors bind at the interfaces between three P-domain spikes when superimposed on the map of the assembled virus (21–23). In GII.4 T = 4 particles, the interface between dimeric P-domain spikes forms around the 20 icosahedral threefold axes between D subunits, and the 60 quasi-threefold axes between A, B, and C subunits (Fig. 2). In the T = 3 particles of GII.2, GI.7, and GI.1, the equivalent interface surrounds the 60 quasi-threefold axes. Thus, a consequence of the larger, T = 4, form is the presence of a larger number of potential receptor attachment sites at the interfaces between three spikes, compared with those in T = 3 particles. On the other extreme, the
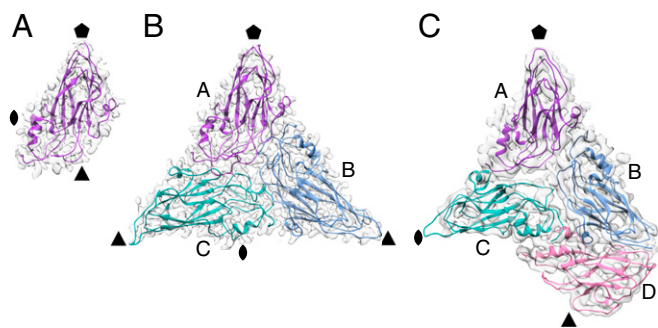
Fig. 3. Asymmetric units of GII.2 T = 1, GII.2 T = 3, and GII.4 T = 4 particles. The icosahedral symmetry axes positions are indicated by a pentagon (fivefold), triangle (threefold), and oval (twofold). (*A*) Shell domain of a single subunit in each ASU of GII.2 T = 1 particles. (*B*) Shell domains of three subunits in quasi-equivalent positions, A, B, and C, that constitute each ASU of GII.2 T = 3 particles. (*C*) GII.4 shell domains in the T = 4 particle ASU with four subunits, A, B, C, and D. The C–D dimer takes up a position similar to the C–C dimer in T = 3 particles.

spikes are placed significantly farther apart (~10–25 Å) in the GII.2 T = 1 particle architecture, and no contacts are made.

Asymmetric Reconstruction Reveals a Zn^{2+} Ion That Affects Shell Stability. Positioned at the outermost surface of the virus, the P2 subdomain bears many of the antigenic features involved in host interactions. Two hypervariable loops (loop A residues 378–381 and loop B residues 295–298) on the outermost apex of P2 in GII.2 are not observed in the existing P-domain crystal structure (11). These loops are only partially resolved in our reconstructions with symmetry imposed. However, subsequent symmetry expansion assigning and extracting 60 icosahedrally related views of the ASU from each whole-particle image, and signal subtraction outside of each ASU, followed by asymmetric focused reconstruction of the ASU enabled near-complete observation of the two missing loops (*Materials and Methods* and *SI Appendix, Fig. S5*).

In HBGA-binding genogroup II strains, the conserved Asp382 (loop A) binds the fucose moiety of human HBGAs (14). In the GII.2 P-domain crystal structure, Asp382 was the last residue observed and the side chain was turned away ~180° from the fucose binding site, which was proposed along with the dynamic loop movement as the possible molecular basis of weak or no HBGA binding by the GII.2 SMV strain. In our cryo-EM structures in the vitrified hydrated state, however, the Asp382 side chain is correctly oriented toward the fucose binding site on all subunits observed in a self-consistent manner (Fig. 4 and *SI Appendix, Fig. S6*).

The two flexible loops appear to be held by a metal-ion density on each subunit. Although partial disorder around the dynamic loop regions limits precise positioning, the metal appears to be bound between the His293 and His299 side chains and a water molecule. The metal-binding site is immediately adjacent to the fucose-binding Asp382 residue. To confirm the identity and

significance of this observed metal binding, we added EDTA, EGTA, or 1,10-phenanthroline to GII.2 VLP samples at the same concentration. The addition of either EDTA or 1,10-phenanthroline that preferentially chelates Zn^{2+} caused an over threefold decrease in the number of intact VLPs observed, and the metal density disappeared in the asymmetric and symmetric reconstructions from the remaining intact particles, whereas EGTA, which preferentially chelates Ca^{2+}, did not cause such changes. The infectivities of murine norovirus and feline caliciviruses that do not bind HBGAs have been reported to be metal dependent (21, 22, 24). Moreover, GII.2 VLPs were reported to bind HBGAs when mixed with GII.2-positive patient stool, underscoring the involvement of coligands (25). This opens a possibility that metal ions may be required for the infectivity and stability of GII.2 SMV and possibly other strains by coordination between the flexible loops as an allosteric coligand, enhancing the human HBGA recognition and virus attachment.

The high-resolution structures of the human outbreak norovirus strain VLPs we have presented here serve as valuable templates for vaccine formulations and development of antivirals that inhibit host attachment and encapsidation. The structural polymorphism between and within genogroups should be valuable in guiding vaccine formulations that cover the broad spectrum of structural differences observed. Careful isolation of specific symmetry forms or a mixed sampling of all forms for vaccine manufacture may help to ensure maximum immunogenicity. The metal ions we identified present the surface antigen loops in the HBGA-binding conformations and allow raising neutralizing immunoglobulins. The addition of metal ions in the vaccine formulations would also be important in ensuring long-term structural integrity of the VLPs during storage of vaccines.

Materials and Methods

VLP Preparation. The GI.1 Norwalk, GII.4 Minerva, and GII.2 Snow Mountain virus strain VLPs were expressed in tobacco leaves. The GI.1 and GII.4 VLPs were prepared commercially by Kentucky Bioprocessing, and the GII.2 VLPs were produced by Hugh S. Mason and Andrew G. Diamos, Arizona State University, Tempe, AZ (26). The GI.7 VLPs were a kind gift from Robert L. Atmar, Baylor College of Medicine, Houston, TX. The VLPs were purified further by size-exclusion chromatography using a Superose 6 Increase 10/300 column (GE) in 20 mM MES·OH (pH 5.75) and 50 mM NaCl, and the VLPs were collected from the void-volume fractions. The samples were concentrated to 3–5 mg/mL using centrifugal concentrators (Amicon).

Cryo-EM Sample Preparation. Lacey carbon grids (300 mesh; Electron Microscopy Sciences) were glow discharged using an easiGlow glow discharger (Pelco). Aliquots (4 μL) of VLP samples were applied to the grids and blotted for 1.7 s using an EM GP2 blotting system (Leica) at 22 °C and 95% humidity level with 10-s preblot incubation time before plunging in liquid ethane for vitrification. For metal-free experiments, the blotting papers were pretreated with 5 mM EDTA and then thoroughly washed in ultrapure water and dried before blotting.

Data Collection. Movies of VLPs embedded in vitreous solution were collected at liquid-nitrogen temperature using a Titan Krios G3 transmission electron microscope (ThermoFisher) equipped with a K2 Summit direct electron detector (Gatan) and a GIF Quantum LS imaging filter (Gatan). The movies were recorded in superresolution mode using EPU acquisition software (ThermoFisher) at 130,000× magnification with a pixel size of 0.535 Å per pixel, later resampled twofold to 1.07 Å per pixel, and nominal defocus range of 1.0–2.8 μm. The total electron exposure of each movie was ~70–85 e/Å^2 with a nominal exposure rate of 2.0 electrons·Å^{-2}·s^{-1} per frame, fractionated into 35 movie frames with 200 ms per frame exposure time.

Data Processing. Many of the software packages used for data processing were available through SBGrid (27). Beam-induced motion of particles was corrected using UNBLUR (*cis*TEM) on whole frames (28, 29). Contrast transfer function (CTF) parameters were estimated from sums of three movie frames using CTFFIND4 (*cis*TEM) (29, 30). Particles were automatically picked ab initio using soft-edged disk templates internally generated in *cis*TEM (29, 31). The picked particle images were boxed, extracted, and 2D-classified ab
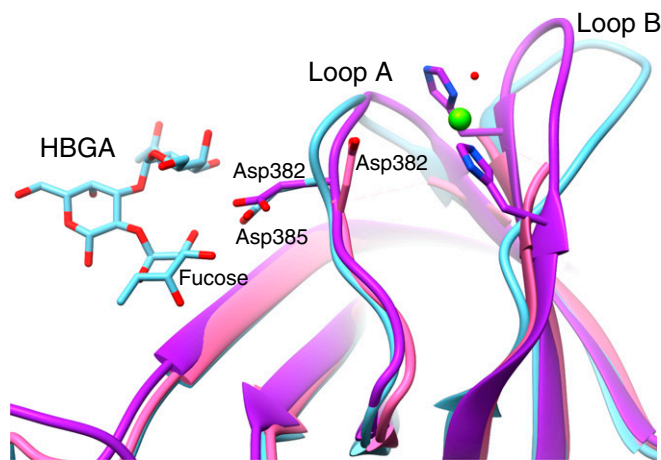

Fig. 4. Histo-blood group antigen binding site in GII.2. The P-domain crystal structure of GII.10 with HBGA bound (PDB ID code 3PA1) was superimposed on the partial P-domain crystal structure of GII.2 (PDB ID code 4RPB) and our GII.2 full-subunit cryo-EM structure without HBGA bound to show the expected positioning of the HBGA binding site of GII.2. In HBGA-binding strains such as GII.10 (light blue), a conserved residue (Asp385) binds the fucose moiety of HBGAs. In the P-domain crystal structure of GII.2 (pink), the side chain of the equivalent residue (Asp382) was turned ~180° away from the fucose binding site, and the hypervariable loop A bearing Asp382 and loop B were not observed. In our cryo-EM structures (purple), the hypervariable loops A and B are resolved with the Asp382 side chain pointed correctly toward the fucose binding site. A Zn^{2+} metal ion (green sphere) is bound between loops A and B immediately adjacent to the fucose binding site that may play a role as a coligand that enhances HBGA binding of the GII.2 strain.

initio into 64 classes using *cis*TEM (*SI Appendix*, Fig. S7) (29). The initial 3D reconstructions were carried out ab initio, followed by 3D refinement using FrealignX (*cis*TEM) with refinement of CTF estimations (*SI Appendix*, Fig. S8) (29, 32). The numbers of movies and particles used toward final reconstructions are in *SI Appendix*, Table S1.

For asymmetric focused reconstruction, a 3D binary mask of an ASU was generated using Chimera and IMAGIC (33, 34). An inverse version of this mask was then applied to the symmetrical 3D reconstruction using apply_mask within *cis*TEM. This masked reconstruction was used to subtract out all but one ASU from the raw images in combination with symmetry expansion. For symmetry expansion, each image was subtracted 60 times using each of the 60 icosahedral ASUs to create a dataset of isolated asymmetric units, which were automatically centered. This dataset of centered isolated asymmetric units was further refined in *cis*TEM without applying symmetry. Symmetry expansion, signal subtraction, and ASU particle image cropping were carried out using the program symmetry_expand_stack_and_par, developed for this project and included with *cis*TEM.

The crystal structure of GI.1 (PDB ID code 1IHM) was docked into the GI.1 cryo-EM densities using MOLREP (CCP-EM) (35, 36) and rebuilt where needed. Water molecules were added using Coot to fit the cryo-EM densities (37). The S-domain structure of GI.1 was initially fitted into the GI.7 EM densities using MOLREP, and manually rebuilt with the correct amino acid sequence for GI.7 using Coot. A P-domain crystal structure of GI.7 (PDB ID code 4P26) was fitted into the EM densities using MOLREP, and modified using Coot to fit the EM densities.

The GII.2 S-domain atomic models were built manually into the EM densities using Coot. The partial crystal structure of GII.2 (PDB ID code 4RPB) was fitted using MOLREP, modified to fit the EM densities, and manually built in previously missing regions using Coot. The manually built GII.2 S-domain model was fitted into the GII.4 EM densities using MOLREP, and manually rebuilt with the correct amino acid sequence for GII.4 using Coot. A crystal structure of the GII.4 P domain (PDB ID code 5IYN) was fitted and modified into the P-domain EM densities using MOLREP and Coot.

All protein models were real space-refined using PHENIX (38), and evaluated using Coot and the MolProbity server (39). Water molecules were added based on identifying EM density peaks with appropriate shape and hydrogen-bond interactions. The symmetric and asymmetric reconstruction cryo-EM maps were deposited in the Electron Microscopy Databank, and the coordinates of the atomic models were deposited in the Protein Data Bank (40, 41). The figures were generated using UCSF Chimera and ChimeraX (33).

Further details are provided in *SI Appendix*.

1. Centers for Disease Control and Prevention (CDC), Norovirus worldwide. https://www.cdc.gov/norovirus/trends-outbreaks/worldwide.html. Accessed 22 May 2019.
2. A. J. Hall *et al.*; Division of Viral Diseases, National Center for Immunization and Respiratory Diseases, Centers for Disease Control and Prevention, Updated norovirus outbreak management and disease prevention guidelines. *MMWR Recomm. Rep.* **60**, 1–18 (2011).
3. M. D. Kirk *et al.*, World Health Organization estimates of the global and regional disease burden of 22 foodborne bacterial, protozoal, and viral diseases, 2010: A data synthesis. *PLoS Med.* **12**, e1001921 (2015).
4. P. F. M. Teunis *et al.*, Norwalk virus: How infectious is it? *J. Med. Virol.* **80**, 1468–1476 (2008).
5. M. de Graaf, J. van Beek, M. P. G. Koopmans, Human norovirus transmission and evolution in a changing world. *Nat. Rev. Microbiol.* **14**, 421–433 (2016).
6. E. F. Donaldson, L. C. Lindesmith, A. D. Lobue, R. S. Baric, Viral shape-shifting: Norovirus evasion of the human immune system. *Nat. Rev. Microbiol.* **8**, 231–241 (2010).
7. B. V. V. Prasad *et al.*, X-ray crystallographic structure of the Norwalk virus capsid. *Science* **286**, 287–290 (1999).
8. S. Vongpunsawad, B. V. Venkataram Prasad, M. K. Estes, Norwalk virus minor capsid protein VP2 associates within the VP1 shell domain. *J. Virol.* **87**, 4818–4825 (2013).
9. M. J. Conley *et al.*, Calicivirus VP2 forms a portal-like assembly following receptor engagement. *Nature* **565**, 377–381 (2019).
10. J. M. Choi, A. M. Hutson, M. K. Estes, B. V. Prasad, Atomic resolution structural characterization of recognition of histo-blood group antigens by Norwalk virus. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 9175–9180 (2008).
11. B. K. Singh, M. M. Leuthold, G. S. Hansman, Structural constraints on human norovirus binding to histo-blood group antigens. *mSphere* **1**, e00049-16 (2016).
12. G. Leroux-Roels *et al.*, Safety and immunogenicity of different formulations of norovirus vaccine candidate in healthy adults: A randomized, controlled, double-blind clinical trial. *J. Infect. Dis.* **217**, 597–607 (2018).
13. U. Katpally, C. E. Wobus, K. Dryden, H. W. Virgin, IV, T. J. Smith, Structure of antibody-neutralized murine norovirus and unexpected differences from viruslike particles. *J. Virol.* **82**, 2079–2088 (2008).
14. B. K. Singh, M. M. Leuthold, G. S. Hansman, Human noroviruses' fondness for histo-blood group antigens. *J. Virol.* **89**, 2024–2040 (2015).
15. R. Pogan, J. Dülfer, C. Uetrecht, Norovirus assembly and stability. *Curr. Opin. Virol.* **31**, 59–65 (2018).
16. G. K. Shoemaker *et al.*, Norwalk virus assembly and stability monitored by mass spectrometry. *Mol. Cell. Proteomics* **9**, 1742–1751 (2010).
17. S. C. Harrison, The familiar and the unexpected in structures of icosahedral viruses. *Curr. Opin. Struct. Biol.* **11**, 195–199 (2001).
18. A. Bertolotti-Ciarlet, L. J. White, R. Chen, B. V. V. Prasad, M. K. Estes, Structural requirements for the assembly of Norwalk virus-like particles. *J. Virol.* **76**, 4044–4055 (2002).
19. S. C. Harrison, A. J. Olson, C. E. Schutt, F. K. Winkler, G. Bricogne, Tomato bushy stunt virus at 2.9 Å resolution. *Nature* **276**, 368–373 (1978).
20. G. S. Hansman *et al.*, Structural basis for broad detection of genogroup II noroviruses by a monoclonal antibody that binds to a site occluded in the viral particle. *J. Virol.* **86**, 3635–3646 (2012).
21. T. Kilic, A. Koromyslova, V. Malak, G. S. Hansman, Atomic structure of the murine norovirus protruding domain and soluble CD300lf receptor complex. *J. Virol.* **92**, e00413-18 (2018).
22. C. A. Nelson *et al.*, Structural basis for murine norovirus engagement of bile acids and the CD300lf receptor. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E9201–E9210 (2018).
23. R. C. Orchard *et al.*, Discovery of a proteinaceous cellular receptor for a norovirus. *Science* **353**, 933–936 (2016).
24. Z. Lu, E. D. Ledgerwood, M. M. Hinchman, R. Dick, J. S. L. Parker, Conserved surface residues on the feline calicivirus (FCV) capsid are essential for interaction with its receptor feline junctional adhesion molecule A (fJAM-A). *J. Virol.* **92**, e00035-18 (2018).
25. P. R. Harrington, J. Vinjé, C. L. Moe, R. S. Baric, Norovirus capture with histo-blood group antigens reveals novel virus-ligand interactions. *J. Virol.* **78**, 3035–3045 (2004).
26. A. G. Diamos, H. S. Mason, High-level expression and enrichment of norovirus virus-like particles in plants using modified geminiviral vectors. *Protein Expr. Purif.* **151**, 86–92 (2018).
27. A. Morin *et al.*, Collaboration gets the most out of software. *eLife* **2**, e01456 (2013).
28. M. G. Campbell *et al.*, Movies of ice-embedded particles enhance resolution in electron cryo-microscopy. *Structure* **20**, 1823–1828 (2012).
29. T. Grant, A. Rohou, N. Grigorieff, *cis*TEM, user-friendly software for single-particle image processing. *eLife* **7**, e35383 (2018).
30. A. Rohou, N. Grigorieff, CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).
31. F. J. Sigworth, Classical detection theory and the cryo-EM particle selection problem. *J. Struct. Biol.* **145**, 111–122 (2004).
32. D. Lyumkis, A. F. Brilot, D. L. Theobald, N. Grigorieff, Likelihood-based classification of cryo-EM images using FREALIGN. *J. Struct. Biol.* **183**, 377–388 (2013).
33. E. F. Pettersen *et al.*, UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
34. M. van Heel *et al.*, "Four-dimensional cryo-electron microscopy at quasi-atomic resolution: IMAGIC 4D" in *International Tables for Crystallography*, E. Arnold, D. M. Himmel, M. G. Rossmann, Eds. (John Wiley & Sons, 2012), vol. F, Crystallography of Biological Macromolecules, pp. 624–628.
35. T. Burnley, C. M. Palmer, M. Winn, Recent developments in the CCP-EM software suite. *Acta Crystallogr. D Struct. Biol.* **73**, 469–477 (2017).
36. A. Vagin, A. Teplyakov, Molecular replacement with MOLREP. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 22–25 (2010).
37. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).
38. P. V. Afonine *et al.*, Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr. D Struct. Biol.* **74**, 531–544 (2018).
39. V. B. Chen *et al.*, MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12–21 (2010).
40. H. M. Berman *et al.*, The Protein Data Bank. *Acta Crystallogr. D Biol. Crystallogr.* **58**, 899–907 (2002).
41. C. L. Lawson *et al.*, EMDataBank unified data resource for 3DEM. *Nucleic Acids Res.* **44**, D396–D403 (2016).

BIOPHYSICS AND COMPUTATIONAL BIOLOGY

High-resolution cryo-EM structures of outbreak strain human norovirus shells reveal size variations

Supporting Information

James Jung[1,2], Timothy Grant[3], Dennis R. Thomas[1], Chris W. Diehnelt[4], Nikolaus Grigorieff[3,5], and Leemor Joshua-Tor[1,2,*]

[1]W. M. Keck Structural Biology Laboratory, Cold Spring Harbor Laboratory, NY 11724, USA.
[2]Howard Hughes Medical Institute, Cold Spring Harbor Laboratory, NY 11724, USA.
[3]Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, VA 20147, USA.
[4]Biodesign Institute Center for Innovations in Medicine, Arizona State University, Tempe, AZ 85281, USA.
[5]RNA Therapeutics Institute and Howard Hughes Medical Institute, University of Massachusetts Medical School, Worcester, MA 01605, USA

*Correspondence:

Leemor Joshua-Tor
leemor@cshl.edu
516-367-8821

Fig. S1. Schematic diagrams of norovirus VP1 major capsid proteins. Norovirus major capsid proteins show a modular organization with a shell domain (S) on the N-terminus forming the icosahedral shell enclosing the genome, and the P-domain on the C-terminus forming dimeric spikes on the virus surface. The P-domain is connected to the S-domain through a long and flexible stalk (T) region. The P-domain consists of two subdomains: P1 emerging from the S-domain and P2, which is an insertion in P1 connected through linkers (L), and positioned at the outermost surface of the virus. The N-terminal arm (NTA) is ordered on the B subunit of only T=3 particles.
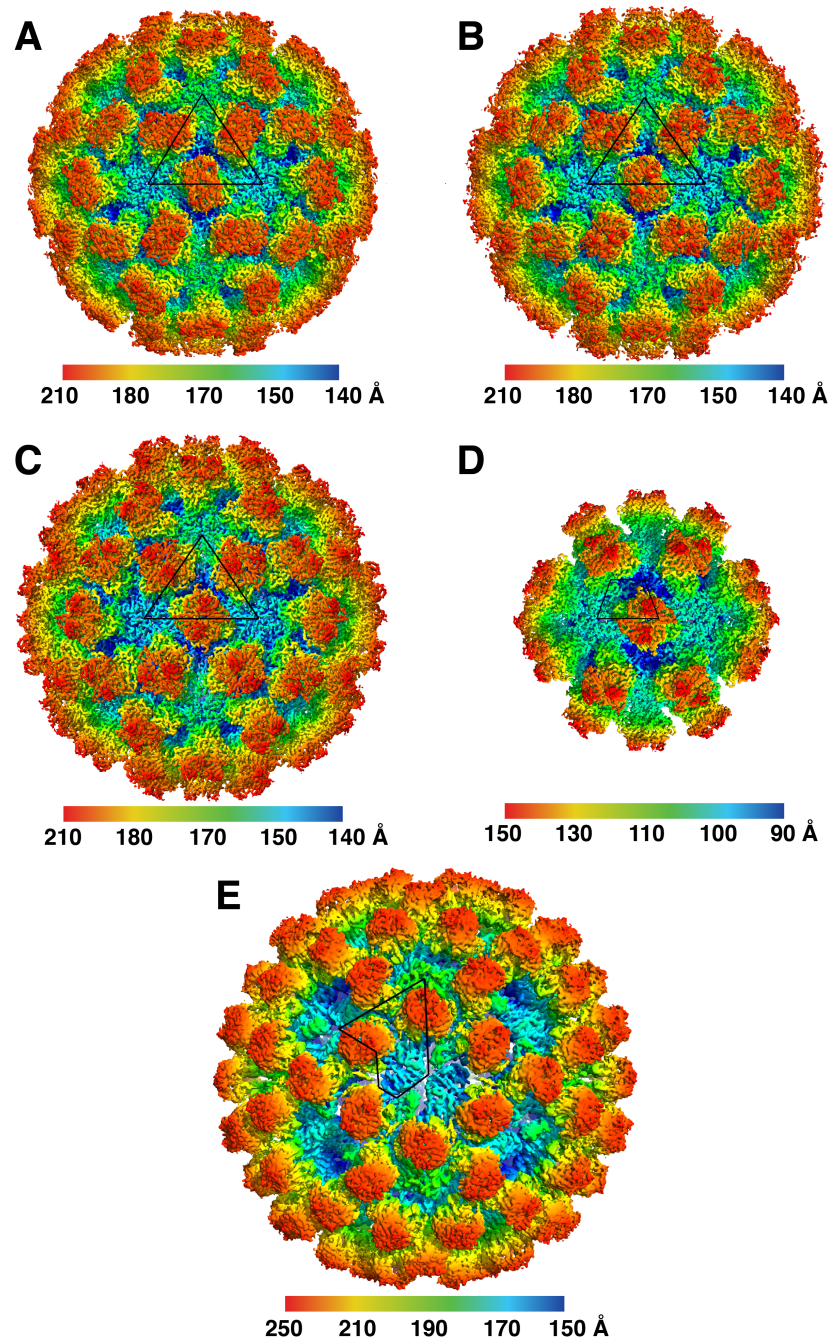
Fig. S2. The cryo-EM structures of human norovirus outbreak strain capsids. Shaded depth-cue representations of norovirus VLP cryo-EM maps viewed along the icosahedral two-fold axis, colored by radial distance from the center (blue to red). The positions of asymmetric units (ASU) are identified by the black lines. (**A**) GI.1 Norwalk strain at 2.9 Å resolution in T=3 icosahedral symmetry with 90 dimeric P-domain spikes assembled from 180 subunits and 410 Å in diameter. (**B**) GI.7 Houston strain at 2.9 Å resolution in T=3 symmetry and 420 Å in diameter. (**C**) GII.2 Snow Mountain Virus strain at 3.1 Å resolution in T=3 symmetry and 430 Å in diameter. The spikes of GII.2 SMV are twisted ~15° counter-clockwise relative to GI.1 and GI.7. (**D**) GII.2 SMV at 2.7 Å resolution in T=1 symmetry with 30 spikes, 60 subunits and 310 Å in diameter. The spikes are placed significantly further apart (~10-25 Å) in T=1 symmetry. (**E**) GII.4 Minerva strain at 4.1 Å resolution in T=4 symmetry with 120 spikes, 240 subunits and 490 Å in diameter. The spikes of GII.4 are twisted ~50° clockwise relative to the orientation of GI.1 and GI.7.
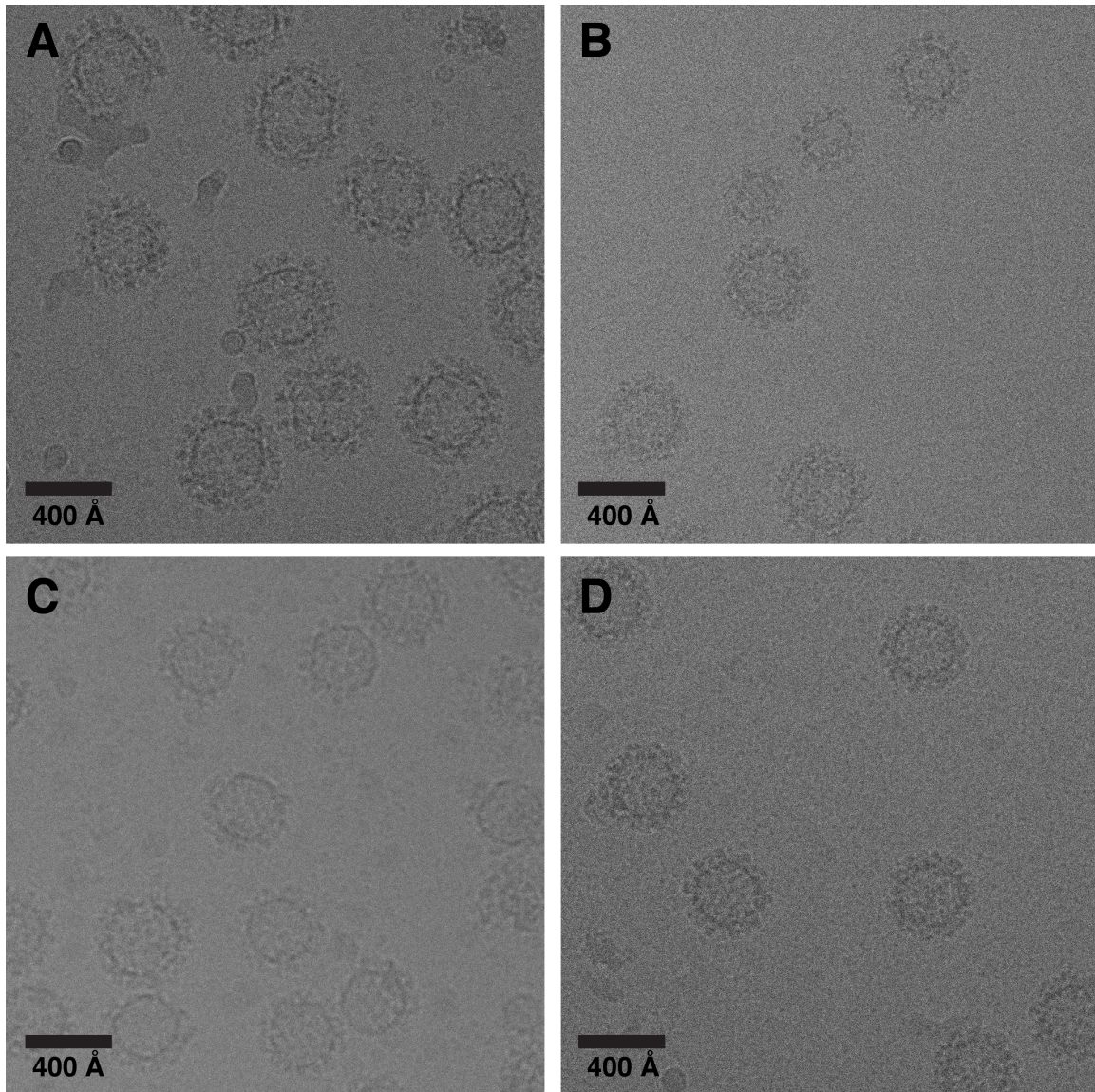
Fig. S3. Representative micrographs of norovirus capsid particles embedded in vitreous ice. (**A**) GII.4 T=4 particles, 490 Å and 280 Å in external and internal diameters, respectively. (**B**) A mixture of both T=3 and T=1 particles of GII.2. The T=3 particles are 430 Å and 240 Å in external and internal diameters, respectively. The T=1 particles are 310 Å and 120 Å in external and internal diameters, respectively. (**C**) GI.7 T=3 particles, 420 Å and 240 Å in external and internal diameters, respectively. (**D**) GI.1 T=3 particles, 410 Å and 240 Å in external and internal diameters, respectively.
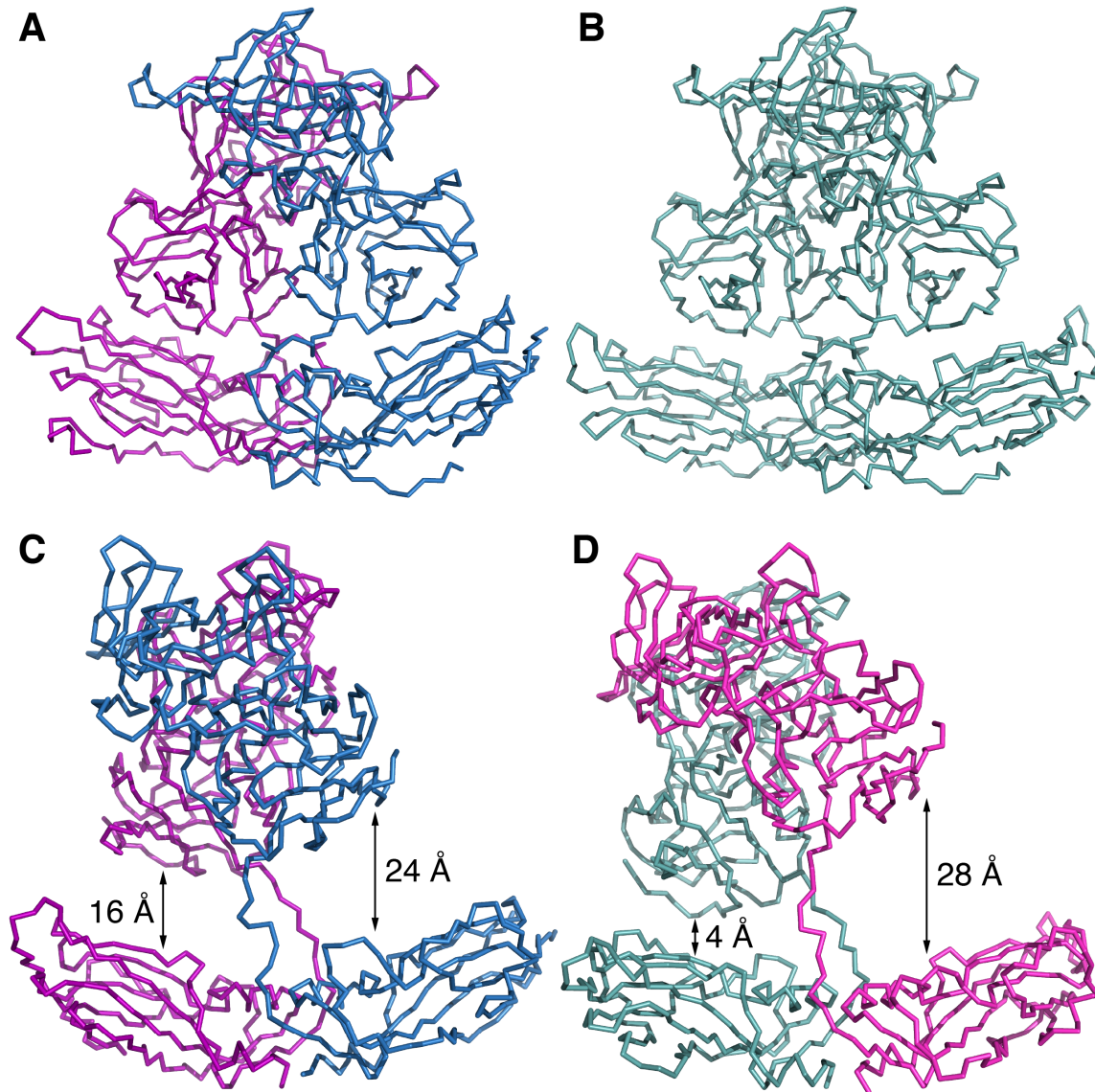
Fig. S4. Bent and flat major capsid dimers constituting norovirus particles. (**A**) The GII.2 T=3 particles are assembled from two types of major capsid dimers, A/B dimers (A in purple and B in blue) with the two S-domains in bent conformation and (**B**) C/C dimers (turquoise) with the S-domains in more flat conformation. The GII.2 T=1 particles are assembled with only bent dimers. (**C**) The GII.4 T=4 particles are assembled from bent A/B (A in purple and B in blue) and (**D**) flat C/D dimers (C in turquoise and D in pink). The P-domains of GII.4 A and C subunits are placed closer to the S-domain (16 and 4 Å, respectively) than B and D subunits (24 and 28 Å, respectively) with the P-domain of the C subunit pulling the S-domain from the shell.
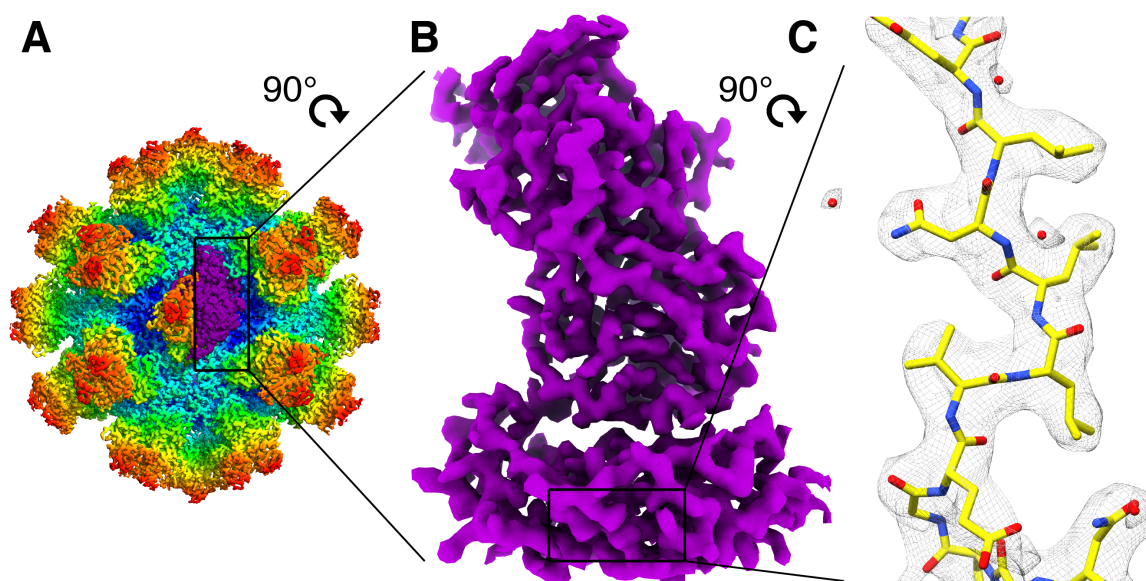
Fig. S5. The asymmetric focused reconstruction of GII.2 asymmetric unit. (**A**) An asymmetric unit (ASU) on the GII.2 T=1 particle, colored purple. (**B**) A shaded depth-cue representation of the cryo-EM map at 2.8 Å resolution in side view. (**C**) A representative section from the S-domain showing the cryo-EM map (gray mesh), fitted atomic model (yellow) and bound water molecules (red spheres).
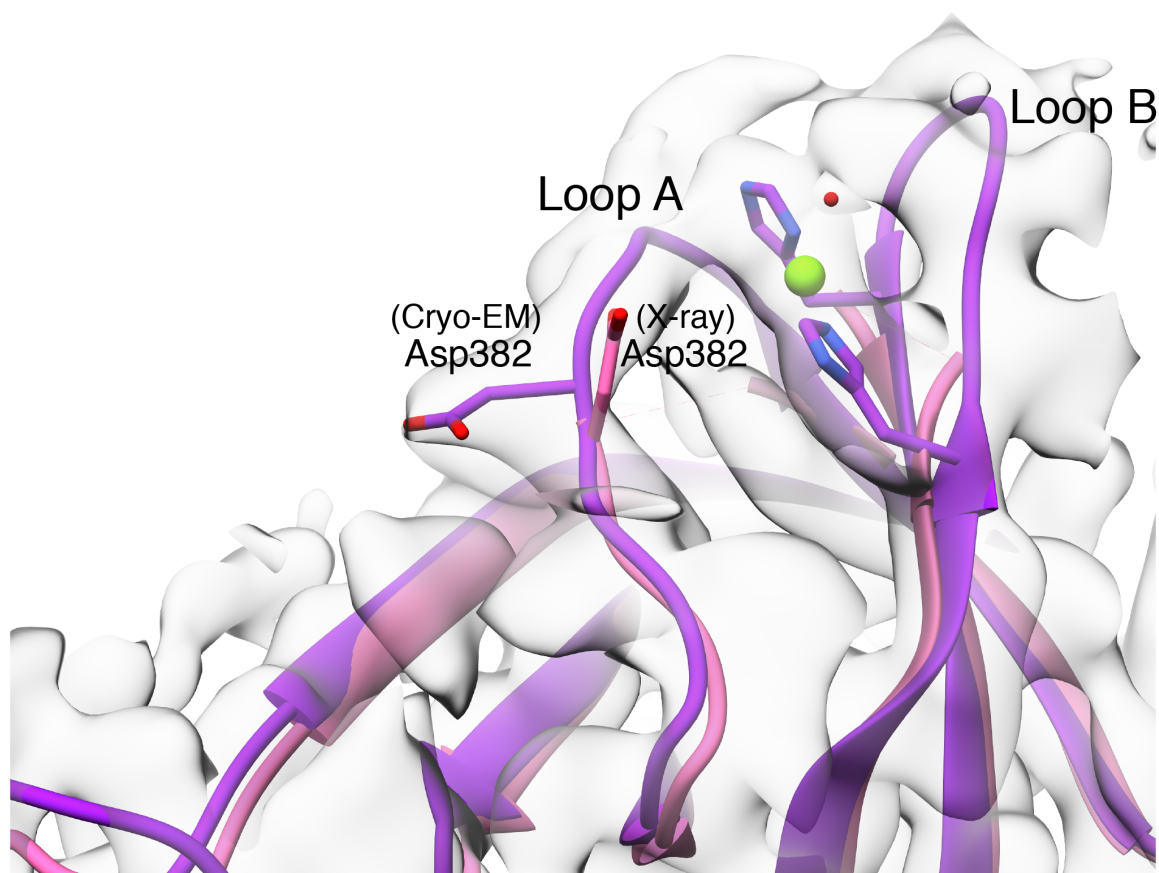
Fig. S6. Asymmetric reconstruction cryo-EM map of the HBGA binding site in GII.2. The hypervariable loops A and B are resolved in the structure of GII.2 reported here, as illustrated clearly by the cryo-EM map, shown in translucent gray along with the structure (purple). The crystal structure of the P-domain of GII.2 (PDB: 4RPB, pink) is superimposed for reference.
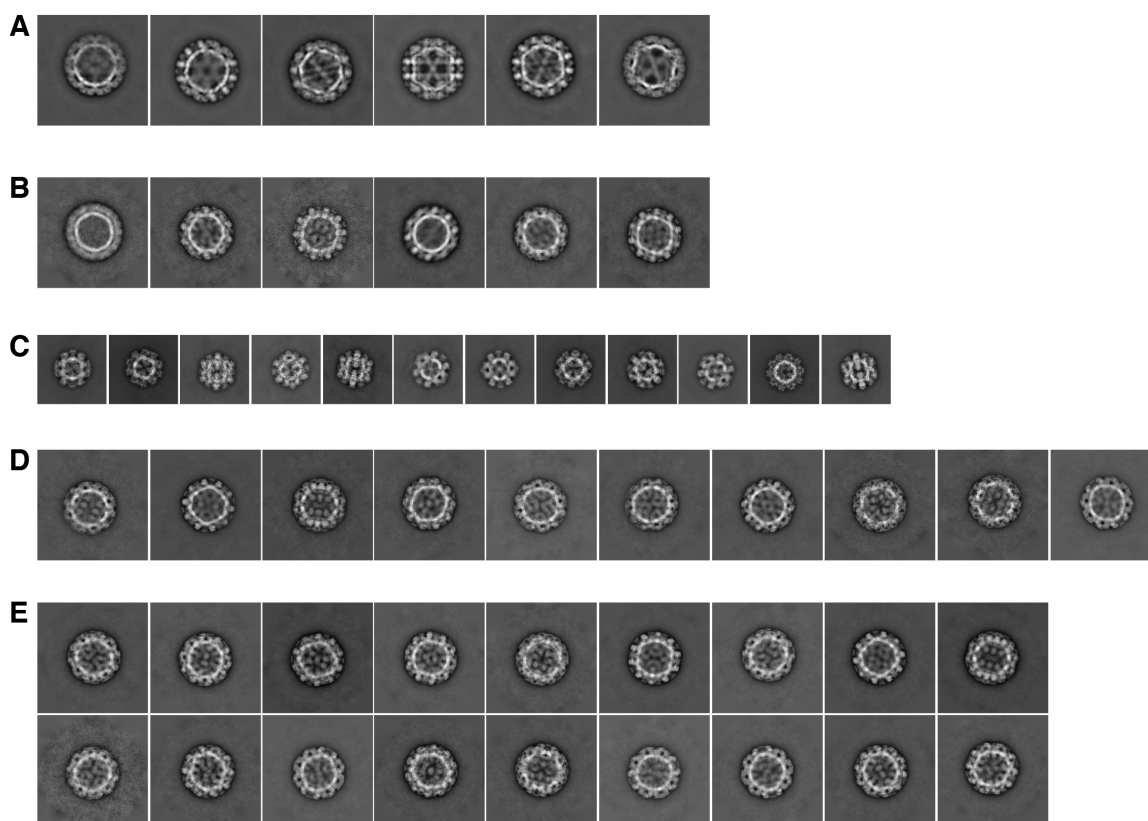
Fig. S7. Cryo-EM 2D class averages. (**A**) GII.4, T=4 particles, (**B**) GII2 T=3 particles, (**C**) GII.2 T=1 particles, (**D**) GI.7 T=3 particles, and (**E**) GI.1 T=3 particles. The box sizes were 500 x 500 Å in **C**, and 800 x 800 Å in **A**, **B**, **D**, and **E**.
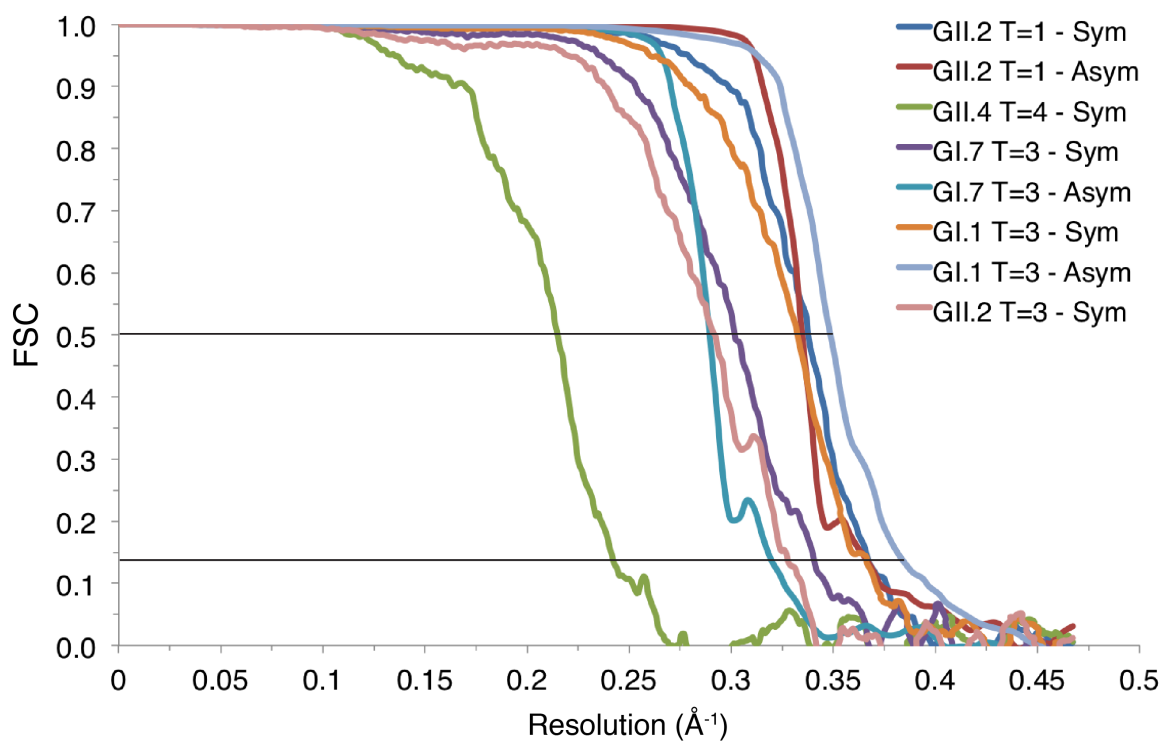
Fig. S8. Fourier shell correlation (FSC) plots. The resolution estimation of the symmetric (Sym) and asymmetric (Asym) reconstructions was based on the 0.143 criterion for the comparison between two half data sets.

Table S1. Cryo-EM data collection, reconstruction and model refinement parameters. The particle numbers used towards the final reconstructions were significantly lower than the initial number of particles picked automatically for GII.2 and GI.7 strains, due to heterogeneity from symmetry variations and broken particles.

GII.4 Minerva, T=4 (PDB: 6OUU, EMDB: EMD-20206)

Number of movies	2,951
Number of particles (total)	94,501
Number of particles (final reconstruction)	17,847
Defocus range (µm)	-1.0 to -2.6
Electron dose (electrons/Å^2)	69.3
Real-space model refinement	
Resolution (Å) / resolution limit[†]	4.1 / 5.7
R_{work} / R_{free}	0.35/0.35
Number of residues	486 (46-530) (residues 1-45 and 531-540 missing)
R.M.S. deviations	
Bond lengths (Å)	0.006
Bond angles (°)	1.080
Ramachandran plot values	
Favored (%)	90.5
Allowed (%)	9.5
Disallowed (%)	0.0

[†]Resolution limit beyond which the Fourier shell correlation (FSC) should indicate an unbiased resolution estimate of the final reconstructions (1, 2).

GII.2 Snow Mountain Virus (SMV), T=3 (PDB: 6OTF, EMDB: EMD-20195)

Number of movies	2,809
Number of particles (total)	40,744
Number of particles (final reconstruction)	1,540
Defocus range (µm)	-1.2 to -2.6
Electron dose (electrons/Å^2)	69.3
Real-space model refinement	
Resolution (Å) / resolution limit[†]	3.1 / 3.9
R_{work} / R_{free}	0.34 / 0.34
Number of residues	521 (12-532) (residues 1-11 and 533-542 missing)
R.M.S. deviations	
Bond lengths (Å)	0.006
Bond angles (°)	0.868
Ramachandran plot values	
Favored (%)	96.1
Allowed (%)	3.9
Disallowed (%)	0.0

[†]Resolution limit used during refinement. The Fourier shell correlation (FSC) curve beyond this limit indicates an unbiased resolution estimate of the final reconstructions (1, 2).

GII.2 SMV, T=1 (PDB: 6OUC, EMDB: EMD-20201 [symmetric], EMD-20202 [asymmetric])

Number of movies	2,809
Number of particles (total)	76,405
Number of particles (final reconstruction)	
Symmetric	24,768
Asymmetric	1,486,080
Defocus range (µm)	-1.2 to -2.6
Electron dose (electrons/Å^2)	69.3
Real-space model refinement	
Resolution (Å) / resolution limit[†]	
Symmetric	2.7 / 3.2
Asymmetric	2.8 / 3.2
R_{work} / R_{free}	0.30/0.30
Number of residues	483 (47-190, 195-533) (residues 1-46, 191-194 and 534-542 missing)
R.M.S. deviations	
Bond lengths (Å)	0.008
Bond angles (°)	1.129
Ramachandran plot values	
Favored (%)	96.5
Allowed (%)	3.5
Disallowed (%)	0.0

[†]Resolution limit beyond which the Fourier shell correlation (FSC) should indicate an unbiased resolution estimate of the final reconstructions (1, 2).

GI.7 Houston, T=3 (PDB: 6OU9, EMDB: EMD-20197 [symmetric], EMD-20198 [asymmetric])

Number of movies	1,973
Number of particles (total)	26,362
Number of particles (final reconstruction)	
Symmetric	7,686
Asymmetric	461,160
Defocus range (µm)	-1.4 to -2.8
Electron dose (electrons/Å^2)	78.4
Real-space model refinement	
Resolution (Å) / resolution limit[†]	
Symmetric	2.9 / 3.7
Asymmetric	3.2 / 3.7
R_{work} / R_{free}	0.28/0.28
Number of residues	520 (13-192, 200-527) (residues 1-12, 193-199 and 528-539 missing)
R.M.S. deviations	
Bond lengths (Å)	0.006
Bond angles (°)	0.745
Ramachandran plot values	
Favored (%)	95.7
Allowed (%)	4.3
Disallowed (%)	0.0

[†]Resolution limit beyond which the Fourier shell correlation (FSC) should indicate an unbiased resolution estimate of the final reconstructions (1, 2).

GI.1 Norwalk, T=3 (PDB: 6OUT, EMDB: EMD-20199 [symmetric], EMD-20205 [asymmetric])	
Number of movies	1,820
Number of particles (total)	38,535
Number of particles (final reconstruction)	
Symmetric	4,893
Asymmetric	293,580
Defocus range (µm)	-1.2 to -2.4
Electron dose (electrons/Å^2)	70.0
Real-space model refinement	
Resolution (Å) / resolution limit[†]	
Symmetric	2.7 / 3.4
Asymmetric	2.6 / 3.1
R_{work} / R_{free}	0.30/0.30
Number of residues	512 (9-520) (residues 1-8 and 521-530 missing)
R.M.S. deviations	
Bond lengths (Å)	0.007
Bond angles (°)	0.921
Ramachandran plot values	
Favored (%)	95.2
Allowed (%)	4.8
Disallowed (%)	0.0

[†]Resolution limit beyond which the Fourier shell correlation (FSC) should indicate an unbiased resolution estimate of the final reconstructions (1, 2).

References

1. Shaikh TR, Hegerl R, & Frank J (2003) An approach to examining model dependence in EM reconstructions using cross-validation. *Journal of Structural Biology* 142(2):301-310.
2. Grant T, Rohou A, & Grigorieff N (2018) CisTEM, user-friendly software for single-particle image processing. *eLife* 7.