# Supporting Information

Zhang et al. 10.1073/pnas.1211702109

## SI Materials and Methods

**Alignment of the Cryo-EM Map Density with the Amino Acid Sequence.** A small computer program was written that searched for possible alignments between the known amino acid sequences derived from the genome with the information in the cryo-EM map. Regions of good density that represented approximately 15-amino-acid polypeptides were reduced to a sequence of numbers from one to three that described the volume of the density (estimated by eye) for each of the residues in the polypeptide. Similarly, the translated genome sequences were translated into a sequence of numbers from one to six that represented the volume of each amino acid. A computer program then searched for good correlations between the experimental sequences and the reduced sequence of every gene in the Sputnik. This program is available at http://bilbo.bio.purdue.edu/~viruswww/Rossmann_home/softwares/xinzheng.php.

**Quasi- and Pseudo-Symmetry.** The original formulation of quasi-symmetry by Caspar and Klug (1) considered a 2D hexagonal array of close packed hexamers with the integers h and k identifying the position of hexamers in the array. The array could be folded into an icosahedron by introducing pentamers at specified values of h and k. They argued that the environment of subunits in the hexamers would have a quasi-equivalent environment as subunits in the pentamer. The number of quasi-equivalent subunits in an icosahedron is T60, where the triangulation number is given by $T = h^2 + hk + k^2$. These structural predictions were obeyed fairly well in early structural determination of RNA plant viruses. However, when the structure of picornaviruses became available it was seen that the three subunits in the icosahedral asymmetric unit (VP1, VP2, and VP3) had a similar "jelly-roll" structure but an almost completely different amino acid sequence. This stretched the meaning of quasi-equivalence from "the same subunit structure with a similar environment" ($T = 3$) to mean "a similar subunit structure with a similar environment" ($P = 3$). The term "pseudo" (P) symmetry was applied to this situation (2). In cowpea mosaic virus (CPMV) the two jelly-roll subunits VP2 and VP3 around the threefold axes remain covalently connected to form a "double jelly-roll" capsid protein. Thus, in CPMV (3) there is a pseudohexamer consisting of three VP2-VP3 double jelly-roll structures around each threefold axis and a VP1 pentamer around each fivefold axis. In Sputnik and many other large dsDNA viruses all hexameric capsomers ("hexons") are constructed of three double jelly-roll major capsid proteins, and the pentameric vertices are constructed of five single jelly-roll subunits (pentons). In much of the literature the "T number" specifies the number of jelly-rolls in the icosahedral asymmetric unit. However, this really should be specified as a P number.

**Paper Spray Mass Spectroscopy to Determine Lipid Content of Sputnik.**
***Background.*** The empirical chemical formula for any organic compound containing only carbon, oxygen, and hydrogen atoms is $C_lO_mH_{2n}$, where l, m, and n are integers. Because the nominal masses of carbon and oxygen are 12 and 16, respectively, the nominal mass of any organic compound with such an empirical formula must be an even number. Substitution of a hydrogen in the compound by an amino group, $NH_2$, changes the empirical formula to $C_lO_mNH_{2n+1}$, which would produce an odd nominal mass because the nominal mass of nitrogen is 14. Thus, a positively charged ion, as required for mass spectroscopic detection, requires the addition of a proton creating an even nominal mass number. The argument is unchanged if sulfur or phosphorus are substituted into the structure. Because the empirical formula of lipids corresponds to the above formula, and given the usual atomic composition of lipids, the mass spectrum of a natural occurring positively charged lipid would have an even nominal $m/z$ value in the 600–900 range.

***Virus sample preparation.*** In addition to the sample of Sputnik virus three additional virus samples of known composition were examined as controls. These were human rhinovirus 14 (HRV14), which does not have a lipid membrane (2), *Paramecium bursaria* chlorella virus 1 (PBCV-1), which has a membrane (4), and virus-like particles of Chikungunya virus (CHKV), an alphavirus that has a lipid membrane (5). Lipid was extracted from the virus samples by adding 35 μL of chloroform to 15 μL of a buffered aqueous solution of virus at a concentration between 0.5 and 3 mg/mL. The aqueous solutions of the virus samples contained different amounts of the following buffer salts: HRV (PBS = 137 mM NaCl, 2.7 mM KCl, 10 mM $Na_2HPO_40.2H_2O$, and 2 mM $KH_2PO_4$); CHKV (150 mM Tris); PBCV-1 (50 mM Tris); and Sputnik virus (PBS). The chloroform-extracted Sputnik virus was checked by cryo-EM. Only fragments of particles could be found, showing that the chloroform had successfully separated the lipid membrane from the virus particles.

***Paper spray mass spectrometry.*** Mass analyses were performed using a Thermo Fisher LTQ mass spectrometer (Thermo Scientific) as previously described (6, 7). Chromatographic cellulose paper grade 31ET, 0.5 mm thick, from Whatman was cut into triangles of ~6-mm width × 10-mm height. Thirty to 40 μL of 2:1 methanol:chloroform was added to the paper, and 2.5 μL of the virus sample was then applied to the paper. The spray voltage was set to 4,500 V.

***Nano-electrospray ionization mass spectrometry.*** Virus sample (2.5 μL) was diluted to 22.5 μL using 2:1 methanol:chloroform. Mass analyses were performed using a Thermo Fisher LTQ mass spectrometer (Thermo Scientific).

The nano-electrospray ionization mass spectrometry (nano ESI-MS) spiking experiment was performed to (*i*) exclude lipid signal suppression induced by buffer salts of the Sputnik sample, and (*ii*) obtain further information about the lower limit of lipid detection in the presence of buffer salts. The virus concentration of the Sputnik sample was approximately 0.5 mg/mL before dilution. Addition of 1.5 ng of porcine brain lipids resulted in a final lipid content of 0.4% (wt/wt) compared with that of the virus. The ESI-MS signals related to the brain lipids could be easily detected, whereas the signal from the Sputnik sample (red arrows are uneven-numbered and do not correspond to lipids). Hence, the Sputnik virus has less (probably much less) than 1% (wt/wt) lipid.

1. Caspar DLD, Klug A (1962) Physical principles in the construction of regular viruses. *Cold Spring Harb Symp Quant Biol* 27:1–24.
2. Rossmann MG, et al. (1985) Structure of a human common cold virus and functional relationship to other picornaviruses. *Nature* 317(6033):145–153.
3. Lin T, et al. (1999) The refined crystal structure of cowpea mosaic virus at 2.8 A resolution. *Virology* 265(1):20–34.
4. Zhang X, et al. (2011) Three-dimensional structure and function of the *Paramecium bursaria* chlorella virus capsid. *Proc Natl Acad Sci USA* 108(36):14837–14842.
5. Akahata W, et al. (2010) A virus-like particle vaccine for epidemic Chikungunya virus protects nonhuman primates against infection. *Nat Med* 16(3):334–338.
6. Wang H, Liu J, Cooks RG, Ouyang Z (2010) Paper spray for direct analysis of complex mixtures using mass spectrometry. *Angew Chem Int Ed Engl* 49(5):877–880.
7. Liu J, et al. (2010) Development, characterization, and application of paper spray ionization. *Anal Chem* 82(6):2463–2471.

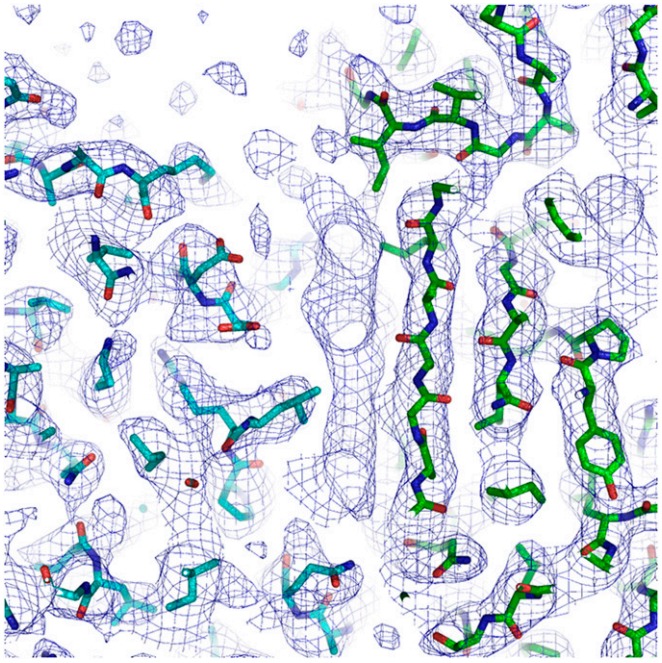

**Fig. S1.** Minor capsid protein. Major capsid protein atomic models from neighboring capsomers colored blue and green were superimposed on the electron density of full Sputnik virus. The uninterpreted density between the two capsomers belongs to the minor capsid protein. The reconstruction of empty virus does not have this density.

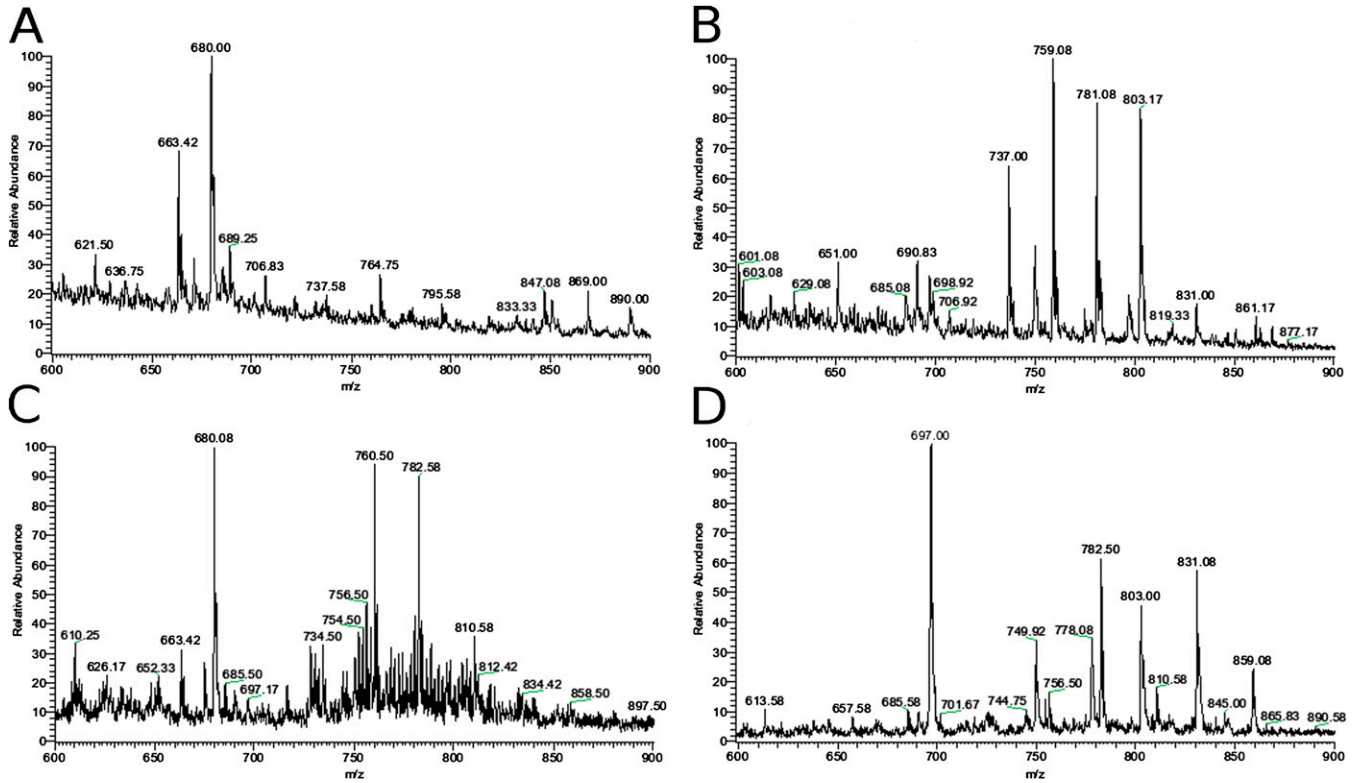


**Fig. S2.** Paper spray mass spectrometry (PS-MS) analyses of (*A*) Sputnik virus, (*B*) HRV14, (*C*) PBCV-1, and (*D*) CHIKV in the positive ion mode. The lipid-containing sample of PBCV-1 shows even *m/z* values (734.5, 754.5, 756.5, 760.5, 782.6, 810.6, 858.5) typical of phosphatidylserines and phosphatidylcholines. The lipid-containing sample of CHIKV also shows even *m/z* values (756.5, 782.6, 810.6, 890.6) typical of phosphatidylcholines. In contrast, the Sputnik and HRV do not show typical, even-numbered, lipid signals.

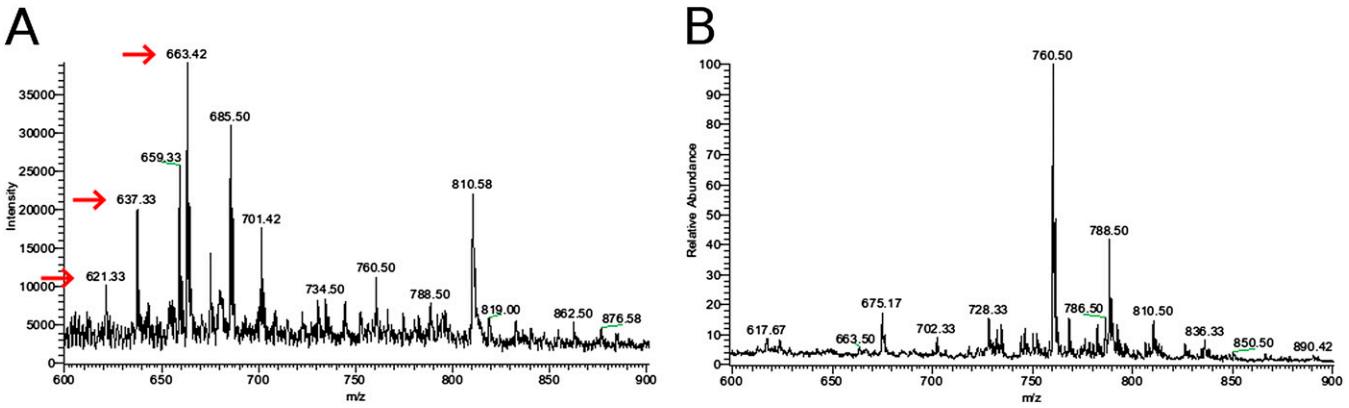


**Fig. S3.** Nano ESI-MS analysis. (*A*) Nano ESI-MS of 2.5 μL of chloroform-extracted material from the Spunik virus with 1.5 ng of polar porcine brain lipids, compared with porcine brain lipids (*B*). Signals related to the Sputnik virus extract are marked with red arrows. A series of typical even-numbered *m/z* values (734.5, 760.5, 788.5, 810.5) are related to the porcine brain lipids. (*B*) PS-MS of an extract of polar porcine brain lipids (the standard solution was obtained from Avanti Polar Lipids) in the positive ion mode. The sample showed a series of typical even-numbered signals.

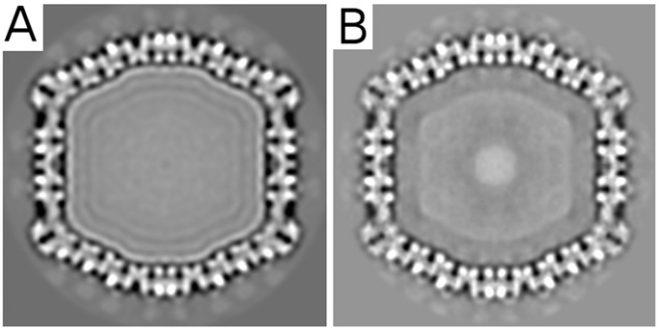


**Fig. S4.** Three density layers underneath the Sputnik capsid. (*A*) Densities of three layers were enhanced by applying a 20-Å low-pass filter. The outermost layer has higher density than the two inner layers. (*B*) Central section of the 20-Å low-pass filtered reconstruction of the empty Sputnik virus. The empty virus does not have any density that might indicate the presence of a membrane.

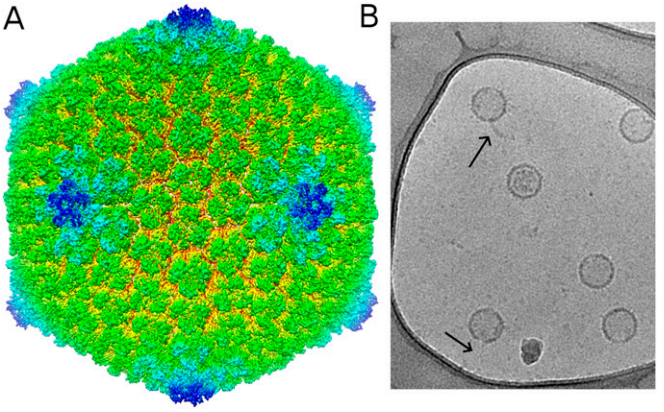


**Fig. S5.** Empty Sputnik virus. (*A*) Surface density of the empty Sputnik virus, colored according to radius. (*B*) A dsDNA-like feature can be seen exiting from two of the particles when the virions were exposed to a pH of 5.5 (arrows).

**Table S1. Potential hydrogen bonds in hexon–hexon interactions**

| Subclass | Neighboring capsomers | | Amino acids pairs having potential hydrogen bonds* |
|---|---|---|---|
| fs-1 | I | II | 1 Asn-315 6 Thr-298; 1 Gln-318 6 Val-316; 1 Asn-36 6 Asp-320; 1 Gln-364 6 Lys-292; 1 Gln-365 6 Arg-295; 1 Asn-438 |
| | Bend | Twist | 6 Gln-318; 1 Asn-440 6 Gln-318; 1 Gln-442 6 Thr-297; 2 Tyr-38 6 Asn-111; 2 Pro-39 6 Gln-49; 2 Gln-266 6 Ser-46; |
| | 8.8 | 2.5 | 2 Gln-266 6 Phe-45 |
| | II | V | 4 Asn-315 12 Gln-299; 4 Gln-318 12 Val-316; 4 Asn-363 12 Ser-294; 4 Asn-363 12 Asp-320; 4 Ser-369 12 Thr-109; |
| | Bend | Twist | 4 Asn-438 12 Gln-318; 4 Asn-440 12 Gln-318; 4 Gln-442 12 Thr-297; 6 Tyr-38 12 Asn-111; 6 Pro-39 12 Gln-49; |
| | 10 | 1.5 | 6 Pro-41 12 Ser-44; 6 Ser-58 12 Gly-116; 6 Ser-58 12 Ser-227; 6 Gln-266 12 Phe-45 |
| | V | IV† | 8 Lys-313 11 Asn-334; 8 Gln-318 11 Val-316; 8 Gln-365 11 Arg-295; 8 Asn-438 11 Asn-438; 8 Asn-438 11 Gln-318; |
| | Bend | Twist | 8 Asn-440 11 Gln-318; 10 Tyr-38 11 Asn-111; 10 Pro-39 11 Asn-51; 10 Pro-41 11 Ser-44; 10 Gln-266 11 Phe-45 |
| | 9.7 | 2.1 | |
| fs-2 | II | III | 7 Asn-115 4 Tyr-38; 7 Thr-226 4 Tyr-38; 7 Asn-315 5 Gln-301; 7 Thr-297 5 Asn-302; 7 Asn-108 5 Ser-309; 7 Asn-438 |
| | Bend | Twist | 5 Gln-318; 7 Ser-480 5 Asn-363; 7 Ile-481 5 Asn-363; 7 Asn-3 5 Asn-363; 7 Ala-479 5 Gln-364; 7 Ala-479 5 Gln-365; |
| | 2.7 | 3.5 | 7 Asp-320 5 Gln-442 |
| | I | V | 2 Asn-115 12 Tyr-38; 2 Gly-116 12 Tyr-38; 2 Lys-292 13 Gln-365; 2 Thr-298 13 Gln-301; 2 Asn-315 13 Gln-301; |
| | Bend | Twist | 2 Asp-320 13 Gln-442; 2 Asn-438 13 Gln-318; 2 Asn-474 12 Pro-57; 2 Thr-475 12 Tyr-38; 2 Ala-479 13 Gln-364; |
| | 13 | 1.7 | 2 Ser-480 13 Asn-363; 2 Ile-481 13 Asn-363; 2 Gly-482 13 Asn-363 |
| ff-1 | V | I† | 1 Asn-48 13 Asn-236; 1 Gln-49 13 Asn-236; 1 Asn-55 13 Asn-315; 1 Asn-108 13 Asn-48; 1 Asn-11 13 Asn-108; |
| | Bend | Twist | 1 Pro-118 13 Asn-120; 1 Asn-438 13 Asn-55; 1 Asn-474 13 Asp-320; 1 Asn-474 13 Gln-318; 1 Thr-475 13 Asp-320; |
| | 22 | 11 | 1 Ile-481 13 Gly-116 |
| | II† | IV† | 8 Asn-48 5 Asn-236; 8 Pro-118 5 Asn-120; 8 Leu-234 5 Asn-236; 8 Asp-320 5 Ser-227; 8 Asp-320 5 Leu-228; 8 Asn-438 |
| | Bend | Twist | 5 Asn-55; 8 Asn-474 5 Asp-320 |
| | 23 | 10.7 | |
| ff-2 | II | IV | 4 Ser-46 9 Thr-297; 4 Asn-48 9 Arg-295; 4 Gln-49 9 Arg-295; 4 Asn-235 9 Asn-108; 4 Asn-236 9 Asn-334; 4 Arg-295 |
| | Bend | Twist | 9 Gln-49; 4 Thr-297 9 Ser-46; 4 Thr-297 9 Ser-44; 4 Gln-299 9 Ser-47; 4 Gln-318 9 Asn-55; 4 Asn-334 9 Asn-48; |
| | 5 | 17 | 4 Asn-334 9 Asn-236; 4 Asn-334 9 Ser-47 |
| ff-3 | IV | IV† | 10 Phe-45 10 Gln-299; 10 Asn-48 10 Tyr-338; 10 Asn-48 10 Asn-235; 10 Gln-299 10 Phe-45; 10 Tyr-338 10 Asn-48; |
| | Bend | Twist | 10 Asn-235 10 Asn-48; 10 Asn-236 10 Asn-48; 10 Asn-48 10 Asn-236; 10 Asn-236 10 Asn-236; |
| | 2.8 | 16 | |
| ss-1 | I | I† | 1 Gln-33 2 Arg-428; 1 Gln-34 2 Gln-364; 1 Thr-36 2 Asn-363; 1 Tyr-38 2 Asn-363; 1 Pro-57 2 Asn-438; 1 Pro-57 |
| | Bend | Twist | 2 Asn-440; 1 Thr-61 2 Asn-363; 1 Phe-273 2 Asn-363; 1 Thr-275 2 Arg-428; 1 Pro-276 2 Arg-428; 3 Gln-318 3 Asn-55; |
| | 30 | 1.1 | 3 Asn-362 3 Ser-58; 3 Asn-363 3 Tyr-38; 3 Asn-363 3 Phe-273; 3 Asn-363 3 Thr-275; 3 Gln-365 3 Tyr-37; 3 Gln-365 |
| | | | 3 Gln-365; 3 Arg-428 3 Gln-33; 3 Asn-440 3 Pro-57 |
| | V | II† | 13 Gln-34 6 Gln-364; 13 Thr-36 6 Asn-363; 13 Tyr-37 6 Gln-365; 13 Pro-57 6 Asn-440; 13 Phe-273 6 Asn-363; |
| | Bend | Twist | 13 Pro-276 6 Arg-428; 11 Asn-363 5 Phe-273; 11 Asn-363 5 Gln-34; 11 Asn-363 5 Tyr-38; 11 Gln-364 5 Gln-34; |
| | 29.2 | 4.1 | 11 Gln-365 5 Tyr-37; 11 Gln-365 5 Gly-370; 11 Arg-428 5 Gln-34; 11 Arg-428 5 Thr-275; 11 Asn-440 5 Pro-57 |
| | IIV | IV | 7 Gln-34 9 Gln-364; 7 Thr-36 9 Asn-363; 7 Tyr-37 9 Gln-365; 7 Tyr-38 9 Asn-363; 7 Thr-61 9 Asn-363; 7 Thr-275 |
| | Bend | Twist | 8 Asn-363; 7 Asn-363 8 Thr-36; 7 Asn-363 8 Tyr-38; 7 Asn-363 8 Phe-273; 7 Gln-364 8 Gln-34; 7 Gln-365 8 Tyr-37; |
| | 21 | 7.1 | 7 Gln-365 8 Thr-36; 7 Gly-370 8 Gln-365 |
| ss-2 | IV | V | 10 Gln-365 11 Thr-36; 10 Gln-365 12 Gly-370; 10 Gln-442 11 Pro-39; 9 Thr-36 12 Gln-365; 9 Pro-39 12 Gln-442; |
| | Bend | Twist | 9 Pro-41 12 Asn-315; 9 Pro-57 12 Asn-363 |
| | −11 | 1.2 | |

*Residues from neighboring hexons are listed when there is a likely hydrogen bond between them. In each pair, every residue is represented by the first number showing which major capsid protein monomer the residue belongs to followed by the residue name and the sequence number. The conserved pairs within any subclass of interactions are highlighted in green.

†Capsomers that are not in the same asymmetric unit as those without a dagger footnote symbol (as indicated by faded black Roman numerals in Fig. 5*A*).