The Three-Dimensional Structure of Canine Parvovirus and Its Functional Implications

JUN TSAO, MICHAEL S. CHAPMAN, MAVIS AGBANDJE, WALTER KELLER, KATHY SMITH, HAO WU, MING LUO,* THOMAS J. SMITH, MICHAEL G. ROSSMANN,† RICHARD W. COMPANS, COLIN R. PARRISH

The three-dimensional atomic structure of a single-stranded DNA virus has been determined. Infectious virions of canine parvovirus contain 60 protein subunits that are predominantly VP-2. The central structural motif of VP-2 has the same topology (an eight-stranded antiparallel β barrel) as has been found in many other icosahedral viruses but represents only about one-third of the capsid protein. There is a 22 angstrom (Å) long protrusion on the threefold axes, a 15 Å deep canyon circulating about each of the five cylindrical structures at the fivefold axes, and a 15 Å deep depression at the twofold axes. By analogy with rhinoviruses, the canyon may be the site of receptor attachment. Residues related to the antigenic properties of the virus are found on the threefold protrusions. Some of the amino termini of VP-2 run to the exterior in full but not empty virions, which is consistent with the observation that some VP-2 polypeptides in full particles can be cleaved by trypsin. Eleven nucleotides are seen in each of 60 symmetry-related pockets on the interior surface of the capsid and together account for 13 percent of the genome.

PARVOVIRUSES ARE THE CAUSE OF A NUMBER OF DISEASES IN man and other animals. Autonomous parvoviruses replicate only in dividing cells and hence have a predilection for either young or unborn animals or for tissues of older animals that contain proliferating cells. The human parvovirus, B19, is the etiological agent for transient aplastic crisis [an acute episode of bone marrow failure (*1*)] and of childhood fifth disease [erythema infectosum (*2*)]. After infection of fetal or neonatal animals, canine parvovirus (CPV) may cause myocarditis leading to death by heart failure (*3*). In older animals both viruses infect lymphoid tissue and reduce the number of circulating lymphocytes (panleukopenia). The proliferating epithelial cells of the small intestines are also infected, causing diarrhea and possibly vomiting. Natural transmission of CPV is primarily by the fecal-oral route.

J. Tsao, M. S. Chapman, M. Agbandje, W. Keller, K. Smith, H. Wu, M. Luo, T. J. Smith, and M. G. Rossmann are in the Department of Biological Sciences, Purdue University, West Lafayette, IN 47907. R. W. Compans is in the Department of Microbiology, University of Alabama, University Station, Birmingham, AL 35294. C. R. Parrish is at the James A. Baker Institute for Animal Health, New York State College of Veterinary Medicine, Cornell University, Ithaca, NY 14853.

*Present address: Department of Microbiology, University of Alabama, University Station, Birmingham, AL 35294.
†To whom correspondence should be addressed.

Canine parvovirus is a remarkable example of the emergence of a new viral pathogen. It was first identified in 1978 and within a few months it had become pandemic (*4*). It is now endemic in all populations of domestic and wild canids that have been examined. Sequence and other data (*5*) indicate that CPV is closely related to feline panleukopenia virus (FPV). Genetic recombination mapping shows that four coding and three noncoding changes within the coat protein region can endow FPV with the ability to replicate in dogs (*6*). Viral infection is initiated by attachment to a cell surface receptor (*7*), probably a glycoprotein with terminal sialic acid residues (*8*). Parvoviruses have a diameter of ~255 Å, a molecular mass of between 5.5×10^3 and 6.2×10^3 kD and contain single-stranded DNA of ~5000 bases (*9*). The number of protein species in infectious virions varies in different parvoviruses. In rat virus–like parvoviruses, such as CPV, there are three types of polypeptides (VP-1, VP-2, and VP-3). The VP-1 and VP-2 are formed by alternative splicing of the messenger RNA from the viral DNA (*10*). The complete sequence of VP-2 (64 to 66 kD) is present in VP-1 (83 to 86 kD), which contains an additional amino-terminal domain. The VP-3 is formed only in full (DNA-containing) capsids by cleavage of 15 to 20 amino acids from the amino terminus of VP-2. Full capsids contain 60 copies of a combination of VP-2, VP-3, and some VP-1 and hence should be icosahedral with $T = 1$ symmetry in the nomenclature of Caspar and Klug (*11*). Propagation of the virus in vitro or in vivo generally produces a substantial amount (~50 percent) of empty particles, which contain no VP-3, mostly VP-2, and a few VP-1 subunits. Trypsin treatment of empty particles does not cleave VP-1 or VP-2. Infectious (full) particles always contain at least a small number of VP-3's. In vitro, additional VP-2's can be cleaved to VP-3–like protein by trypsin treatment of full particles (*12, 13*). This result suggests that the amino-terminal portion of VP-2 is internal in empty particles but exposed in full particles. The amino end of VP-1 is unusually basic and, together with polyamines (*14*), may help neutralize the DNA.

The packaged single-stranded DNA is the complement of the coding sense in CPV. In addition to the structural proteins, the genome also codes for one nonstructural protein, NS-1, and possibly for a second, NS-2. NS-1 is covalently linked to the 5′ end of newly synthesized DNA and, when packaged, is external to the viral capsid (*15*). However, in later stages of infection this linkage is often disrupted by extracellular nucleases. The total genome has been sequenced for a number of autonomous and adeno-associated parvoviruses (*16*), including CPV (*6, 17, 18*). A glycine-rich sequence (GSGNGSGGGGGGSGGVG) (*19*), representing residues 22 to 39 in VP-2 of CPV, is well conserved in autonomous parvoviruses.

Table 1. Data collection for the monoclinic form.

Parameter	Full particles	Empty particles	K_2PtBr_6 derivatives of full particles
	Data processed		
Films (no.)	253	47	7
Observed reflections (no.)	970,770	113,209	36,649
Unique reflections (no.)	583,747	104,046	35,089
Internal consistency R_I* (percent)	12.2	15.5	15.3
Accepted reflections greater than $n\sigma(I_h)$* where n is	2	3	2
Difference R_F† (percent)		20.8	31.7
	Percentage of theoretically possible data		
Resolution shell (Å)			
∞–30	72	5	3
30–15	95	18	5
15–10	94	20	5
10–7.5	91	20	5
7.5–5.0	79	17	5
5.0–3.5	61	9	4
3.5–3.0	38	4	2
3.0–2.75	14	2	1

*$R_I = \frac{\Sigma_h \Sigma_i |(\langle I_h \rangle - I_{hi})|}{\Sigma_h \Sigma_i \langle I_h \rangle} \times 100$, where $\langle I_h \rangle$ is the mean of the I_{hi} reflections observed with Miller indices h, calculated for reflections with $I_h > n\sigma(I_h)$.

†$R_F = \frac{\Sigma_h \Sigma_i |(F_{h1} - F_{h2})|}{\Sigma_h \Sigma_i \langle F_h \rangle} \times 100$ where F_{h1} and F_{h2} are the structure amplitudes of the reflection h.

Previous structural investigation of Kilham's rat virus (closely related to CPV) by small-angle neutron scattering (*20*) suggested that the virus consisted of two concentric shells of protein. The outer shell was between 91 and 121 Å in radius, and the inner shell (with less scattering power) was between 58 and 91 Å, which, in light of the results reported here, is a reasonable model. Earlier results from our laboratory (*21*) had shown that CPV could be crystallized into the monoclinic space group $P2_1$ with cell dimensions a = 263.1, b = 348.9, c = 267.2 Å, and β = 90.82°. These crystals diffracted to at least 2.8 Å resolution and contained one virion per crystallographic asymmetric unit. If the particle was assumed to be spherical [a reasonable assumption in light of the appearance of electron micrographs (*22*)], packing considerations would suggest that the particles had an external radius of 129 Å. In this article we report our initial results on the atomic-resolution structure of full and empty CPV derived from these monoclinic crystals.

Structure determination. X-ray diffraction data for various morphologically indistinguishable crystal forms were collected for full (*23*) particles and some isomorphous heavy-atom derivatives of these at the Cornell High Energy Synchrotron Source (CHESS). Data for the empty particles were collected at CHESS and the Synchrotron Radiation Source in Daresbury, England. Oscillation angles varied from 0.3 to 0.7°. In addition, attempts were made to collect very low resolution data by increasing the crystal-to-film distance from 100 to 300 mm to help with the initial phase determination. Crystals were randomly oriented. Initial indexing was usually achieved by the auto-indexing procedure of Kim (*24*), but a small number of difficult films were indexed with the use of INTEX, an interactive graphical program (*25*). Refinement of crystal orientation, intensity measurements, scaling, postrefinement, and averaging were accomplished with the Purdue processing program package (*26, 27*). As by far the largest amount of data was available for the full particles crystallized in the monoclinic space group, structure determination centered on this data set. A summary of the data used in this determination is given in Table 1.

The orientation of the virion, within one asymmetric unit of the crystal unit cell, was obtained from the rotation function (*28*) at a point when only ~35 percent of the data was available. At a later stage the particle orientation was determined more precisely by using the more complete data set described above. The orientations of the two-, three-, and fivefold axes (derived from 4.3 Å data) were fitted to those in a standard icosahedron by a least-squares procedure (*29*) with an error of only 0.05°, corresponding to an error of 1.0 Å at a radius of 130 Å. This established that paravoviruses do indeed possess near perfect icosahedral symmetry. One of the icosahedral twofold axes was only 2.6° inclined to the monoclinic 2_1 axis. Hence, the rotation function peaks arising from the two particles in the cell were almost coincident when 12 Å resolution data were used and could only be well resolved at resolutions better than 7 Å. Packing considerations of spheres into the crystal cell showed that the center of the virus would be near (0.25, y, 0.25). More accurate values were determined by several methods, all agreeing to within 1.5 Å. This included inspection of the Harker section of the native Patterson map, which showed a large peak; a consequence of one of the noncrystallographic, icosahedral axes being almost parallel to the monoclinic 2_1 axis. However, the accuracy of determining the particle position from this Harker peak is compromised by the inclination between the icosahedral and monoclinic axes.

An initial low-resolution model was constructed that consisted of a hollow shell with an outer radius of 128 Å and an inner radius of 85 Å. The outer and inner shell radii were determined by fitting spherical shells to low-resolution x-ray data and from low-angle solution scattering data. However, in the absence of any useful diffraction data with spacing less than 50 Å, the number of nodes in the mean spherical diffraction function remained ambiguous. Thus, an outer radius of 122 Å was also possible.

The spherical shell model with an outer radius of 128 Å was

Table 2. Parts of the structure that have less order (*19*).

Residues	Amino acid sequence	Density	Position
1–39	This includes the conserved sequence (21)TGSGNGSGGGGGGGSGGGVG(39) and the sequence (16)RNER(19) accessible to trypsin in full capsids.	Some of the carboxyl end of this sequence may be visible and partially ordered in the center of the cylindrical structure about the fivefold axes.	Associated with DNA or partly emerging from virus along the fivefold axes in full particles.
93,94	NG	Small but visible.	On threefold spikes.
155–162	ESATQPPT	Low but interpretable.	The tip of the antiparallel ribbons in the cylindrical structures surrounding the fivefold axes.
221–227	SHTGTSG	Low but interpretable.	On threefold spikes.
362–371	GGAQTDENQA	Only the main chain is visible at a low level of density.	Near the base of the threefold spikes in the depression around the twofold axes.

placed into position and used to compute a set of structure factors to 20 Å resolution. The corresponding model phases were applied to the observed structure amplitudes for computation of an electron density map. This map was icosahedrally averaged and Fourier back-transformed to improve the phases and to allow phase extension (*30, 31*). Phases were extended one reciprocal lattice point at a time. This procedure worked well initially, but when 9 Å resolution had been reached (after 21 steps of phase extension) the agreement between observed and calculated structure amplitudes (F_o and F_c) was sufficiently poor that there was considerable concern about the validity of the whole process.

A survey of heavy-atom derivatives was initiated at this juncture in which very partial data (Table 1) were collected for three different compounds. A 9 Å resolution difference map for the K_2PtBr_6 derivative, in which ~5 percent of the theoretically observable reflections and phases obtained as described above were used, showed a very large negative hole when 60-fold averaged, with a depth more than twice any other hole in the map. It was therefore immediately clear that a heavy-atom position had been found but that the phases represented the Babinet opposite solution of the structure (positive density is negative and vice versa due to a phase shift of 180°). The Babinet inversion was due to the assumption of an outer radius of 128 Å instead of 122 Å. The Pt site was refined to the position $X = 19.5$ Å, $Y = -19.0$ Å, and $Z = 129$ Å with respect to the standard orthogonal axes. An electron density map based on single isomorphous replacement (SIR) Pt phases showed a spherical envelope of protein density, further confirming the essential correctness of the earlier phase solution. Knowledge of the heavy-atom site permitted refinement not only of the heavy-atom parameters but also of the particle position by searching for the closest superposition of the symmetry-related Pt difference peaks. The SIR phases to 8 Å resolution, in which the improved particle position was used, were then used for phase initiation and phase extension and resulted in an immediate radical improvement of the agreement between F_o and F_c. Phase extension could then proceed with confidence. The position of the particle center was further refined periodically as phase extension progressed, moving it a total of 1.5 Å to $x = 0.2534$ and $z = 0.2470$. The averaging was

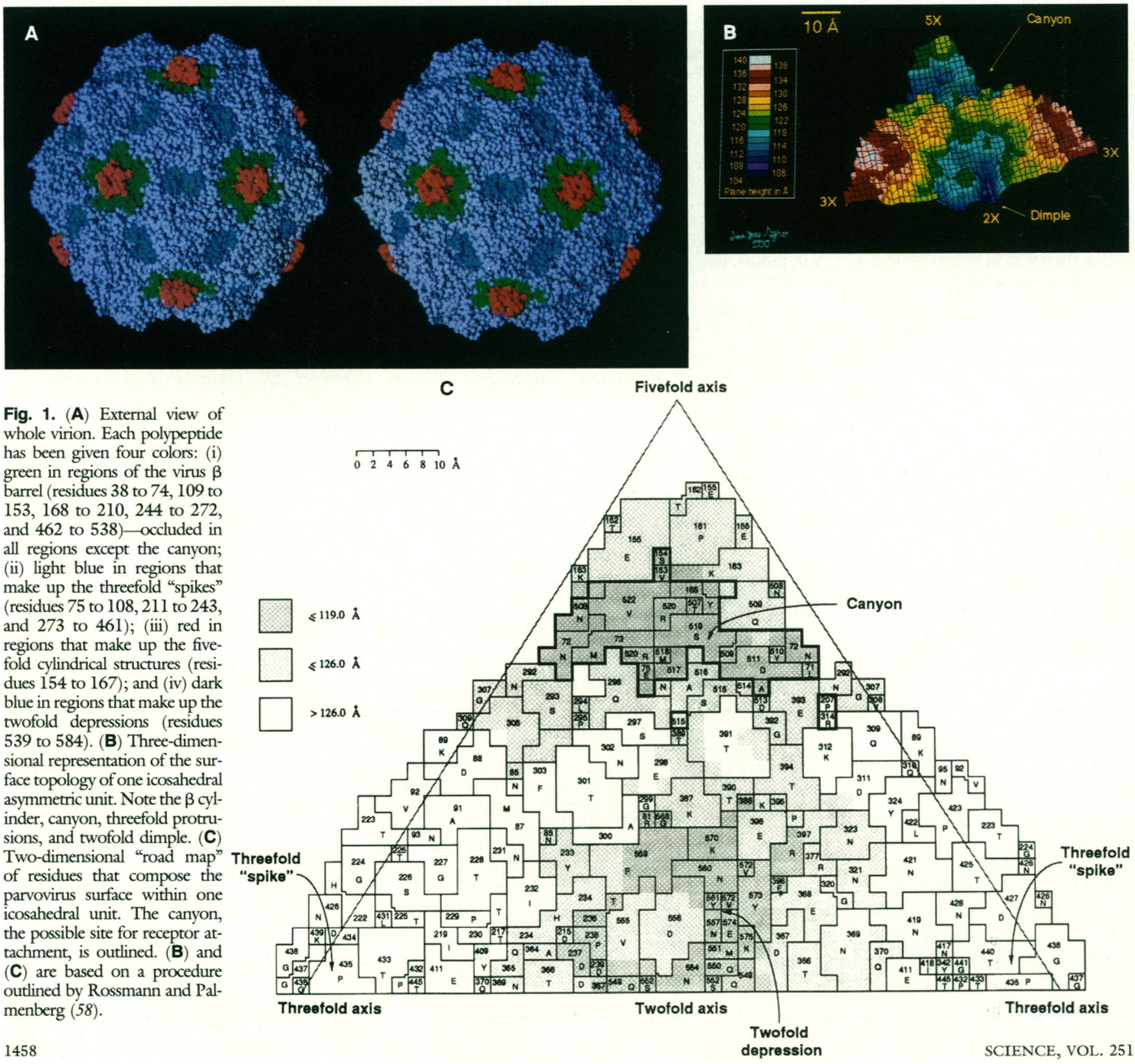

Fig. 1. (**A**) External view of whole virion. Each polypeptide has been given four colors: (i) green in regions of the virus β barrel (residues 38 to 74, 109 to 153, 168 to 210, 244 to 272, and 462 to 538)—occluded in all regions except the canyon; (ii) light blue in regions that make up the threefold "spikes" (residues 75 to 108, 211 to 243, and 273 to 461); (iii) red in regions that make up the fivefold cylindrical structures (residues 154 to 167); and (iv) dark blue in regions that make up the twofold depressions (residues 539 to 584). (**B**) Three-dimensional representation of the surface topology of one icosahedral asymmetric unit. Note the β cylinder, canyon, threefold protrusions, and twofold dimple. (**C**) Two-dimensional "road map" of residues that compose the parvovirus surface within one icosahedral unit. The canyon, the possible site for receptor attachment, is outlined. (**B**) and (**C**) are based on a procedure outlined by Rossmann and Palmenberg (*58*).

performed within an envelope defined by an outer radius of 135 Å and an inner radius of 60 Å until 6 Å resolution phases had been obtained. At this point the outer and inner radii were changed to 140 and 70 Å, respectively. Where there was overlap of envelopes, a tangent plane, equidistant from the overlapping envelope centers, was used to separate adjacent particles. Density outside the envelope (solvent) and inside the inner radius (DNA) was set to zero on every cycle of density modification. An electron density map was computed when phase extension had reached 3.6 Å resolution after 53 steps of phase extension from the SIR Pt phases at 8 Å resolution that were available for 5 percent of the data. Three cycles of phase improvement were performed at each step. The 3.6 Å resolution map was interpreted, but phases were then further extended to 3.25 Å resolution in seven additional steps (*32*).

The averaged electron density was calculated on a 1 Å grid with respect to orthogonal axes of the standard icosahedral setting. The 3.6 Å resolution map was printed on clear plastic sheets and stacked for the initial polypeptide chain tracing and amino acid identification. The twist between strands in the β sheets, the hand of α helices and the chirality of α-carbons showed that the map was a mirror image of the structure. This is the consequence of the arbitrary choice of particle orientation in conjunction with the selected particle center. Thus, the structure was fitted to the mirror image of the 3.25 Å resolution electron density with the interactive graphics program, FRODO (*33*). Unrefined coordinates have been deposited with the Brookhaven Protein Data Bank.

The structure of the protein coat. There are a number of remarkable features on the viral surface (Figs. 1 and 2). (i) There is a 22 Å long, 70 Å diameter "spike" on each of the threefold axes. (ii) A cylindrical structure is formed by five antiparallel β strands forming a sheet about each fivefold axis. The antiparallel strands within a single polypeptide are hydrogen bonded to each other, whereas the five sheets are only loosely associated about each fivefold axis. This cylindrical structure has an internal diameter of between 14 Å at a viral radius of 113 Å to 22 Å at a viral radius of 127 Å. (iii) There is a depression on the twofold axes between the threefold axial "spikes." (iv) A canyon, ~11 Å wide, circulates about each of the fivefold axes immediately outside the cylindrical structure. The canyon floor is ~9 Å below the top of the cylindrical structure on one side and 15 Å below a ridge dividing the canyon from the depression on the twofold axis. These features were not anticipated in light of the featureless appearance of the virus on electron micrographs (*22*). The viral radius is 140 Å along the threefold axes, 132 Å along the fivefold axes, and 112 Å along the twofold axes. The canyon floor is at a radius of 117 Å.

The polypeptide is not visible or barely visible from the amino terminus to residue 37 (*34*) but could be followed without any serious difficulty from residue 38 to the carboxyl terminus at residue 584. The known sequence of CPV–VP-2 was easily associated with the electron density. Sequences that were recognized early in the interpretation were Phe-Trp-Trp (residues 527 to 529) and Tyr-Phe-Gln-Trp (residues 211 to 214). The K_2PtBr_6 heavy-atom compound used in the structure determination was found to bind to Met^{87}, a common mode of binding for Pt compounds (*35*). A mercurial derivative was subsequently shown to have heavy-atom binding sites that were associated with the sulfhydryl groups at 270 and 273. The other two cysteines, at positions 490 and 494, form a disulfide bond.

Residue 38 is near the interior of the capsid close to the fivefold axis. There is a possibility that a few of the amino termini run along the fivefold axis to the exterior of the virus (see section on the cylindrical structure), while the remainder are associated with the DNA. It is indeed quite usual for amino ends of icosahedral viral proteins to be disordered and presumably to be associated with the internal nucleic acid. This is undoubtedly true for the basic amino termini of VP-1 of which there are but a few copies. It is surprising that the carboxyl terminus is not on the viral surface (as is usual in many other icosahedral viruses), although it is on the surface of the isolated protein subunit. Thus, the polypeptide chain must fold on its own completely before assembly of the virion takes place. Parts of the polypeptide (Table 2) that are less ordered are predominantly on the viral surface and include residues near the top of the threefold "spikes" and at the outside end of the cylindrical structures about the fivefold axes.

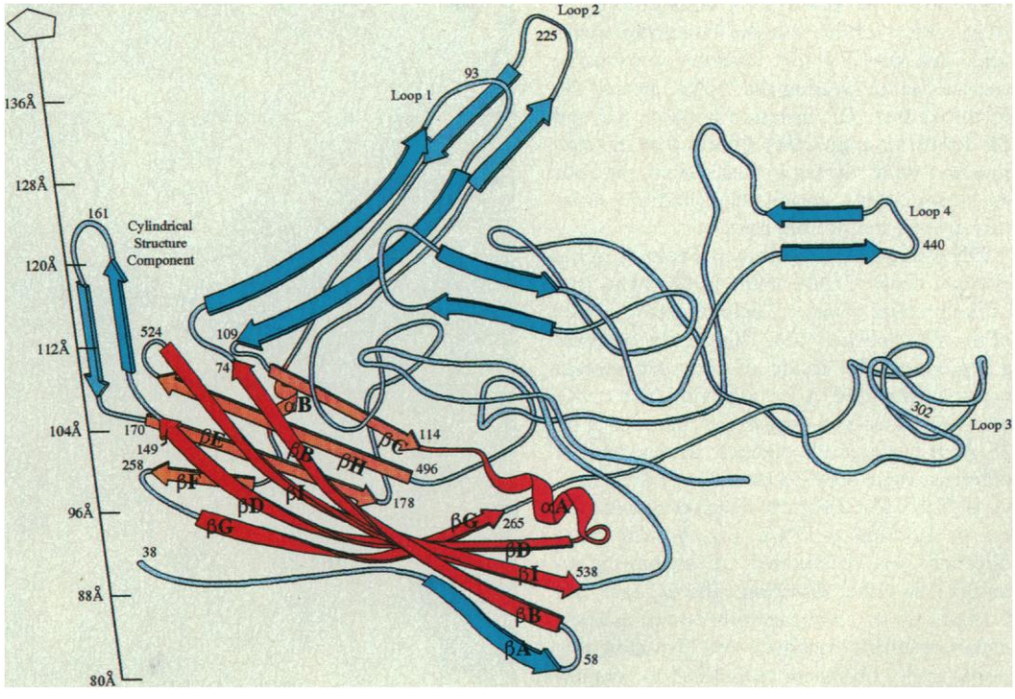

Fig. 2. Ribbon drawing of the VP-2 polypeptide topology. The β-sheet elements that correspond to equivalent structural elements in other viruses are labeled βB, βC . . .βI and are colored red. The large insertions between the β strands of the eight-stranded antiparallel β barrel are blue. The four loops, which together with equivalent structure from two other threefold related subunits make up the threefold "spikes," are labeled loops 1, 2, 3, and 4 (see Table 4). Amino acid sequence numbers are given in strategic locations. The approximate radial distance from the viral center in angstroms is shown along the fivefold axis. The size of the insertions (blue) within the β barrel are βB to βC ≃ 34 residues (loop 1); βC to βD ≃ 15 residues (αA); βD to βE ≃ 20 residues (fivefold cylindrical structure); βE to βF ≃ 99 residues (loop 2 and αβ); βF to βG ≃ 1 residue; βG to βH ≃ 231 residues (loops 3 and 4); and βH to βI ≃ 14 residues.

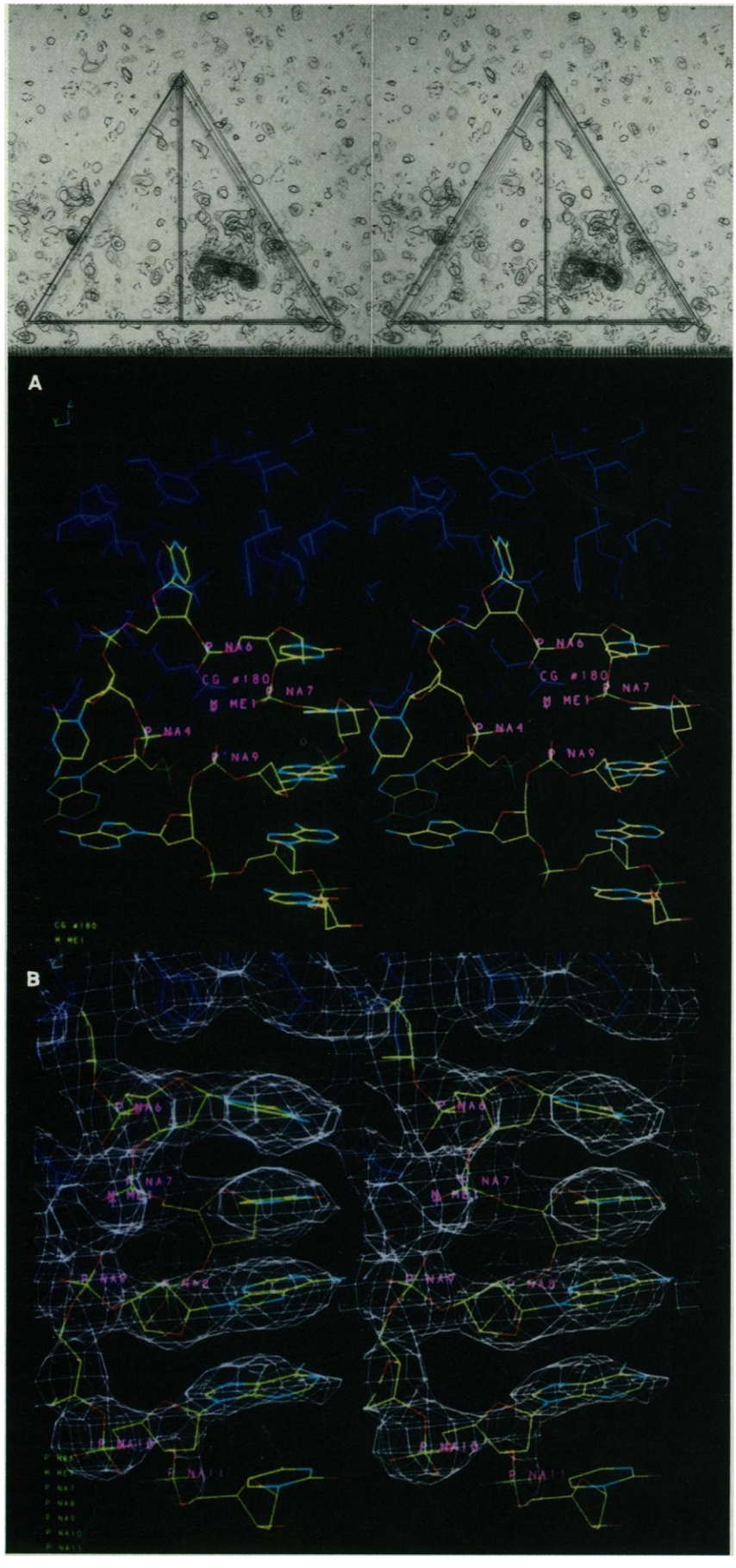

Fig. 3 (top). Photograph of the difference (full – empty virus) electron density map corresponding precisely to features not interpreted as VP-2 polypeptide in the map of the full virus. **Fig. 4 (middle and bottom).** Fit of DNA structure to electron density. The DNA is colored according to atom type; the surrounding protein is shown in dark blue. (**A**) Stacked bases N6, NA7, NA8, and NA10. (**B**) Environment of one of the metal sites in the DNA structure.

The topology of VP-2 (Fig. 2) is remarkably similar to the antiparallel eight-stranded β barrel of other RNA and DNA viruses. The individual β strands along the polypeptide are conventionally named B, C, D, E, F, G, H, and I (*36*). These form two β sheets, BIDG and CHEF, that face each other to form the β barrel. The polypeptide folds up as an antiparallel sheet with strands B, C, D, and E hydrogen-bonded to strands I, H, G, and F, respectively. In many virus coat protein structures there is an α helix (αA) that is inserted between βC and βD as well as an α helix (αB) between βE and βF. These helices also exist in CPV. The virus β-barrel structure motif is situated somewhat internally between a radius of about 82 and 104 Å. Large insertions between the β strands form the viral surface. Differences between viruses are primarily due to the insertions between strands and to the packing of the protein subunits within the icosahedral shell. There is an enormous insertion of CPV between βG and βH, consisting of about 219 residues, that corresponds to the antigenic "FMDV loop" in picornaviruses (*31, 37*) and the major surface feature in plant SBMV (*38*). The GH insertion produces a surface feature near the $T = 3$ pseudo (picornaviruses) or quasi-threefold (SBMV) axes. Similarly, the large insertion in the $T = 1$ CPV virus produces a large surface feature near the icosahedral threefold axes. In CPV this insertion is primarily responsible for creating the "spike" around the threefold axes. The insertion between βD and βE forms an antiparallel ribbon that, in conjunction with the same sheets from the four adjacent proteins, creates the cylindrical structure around the fivefold axes.

DNA structure. A substantial volume of the electron density, comparable in height to that of the protein, could not be interpreted as part of the VP-2 polypeptide. This density occupies a region on the inside of the coat protein between 79 and 99 Å from the viral center with a maximal cross-sectional area of about 25 Å by 25 Å. It corresponds precisely to density in a difference map between the full and the empty virus (Table 1 and Fig. 3). Much of this density can be modeled as DNA (Fig. 4). The chain direction was determined by attempting to model it in either direction. The C_2'-*endo* and C_3'-*endo* ribose conformations as well as commonly occurring backbone and glycosidic torsional angles (*39*) were considered for obtain-

Table 3. Summary of comparison of virus structures. The comparison of T4 phage lysozyme (T4L) and hen egg white lysozyme (HEW) is given as a standard or "benchmark."

Molecule 1	Molecule 2	Amino acids in molecule 1* (no.)	Amino acids in molecule 2* (no.)	Equivalent residues (no.)	Root-mean-square deviation of Cα atoms (Å)	Topologically equivalent residues (percent)	
						Molecule 1	Molecule 2
CPV	HRV14-VP1	469	172	99	3.7	21	58
CPV	HRV14-VP2	469	183	114	3.9	24	62
CPV	HRV14-VP3	469	162	79	3.6	17	42
CPV	SBMV (subunit A)	473	177	112	3.8	24	63
CPV	STNV	472	162	87	3.7	18	53
T4L	HEW	164	129	78	4.8	48	60

*The number of residues given are those between the beginning of βB and the end of βI.

Table 4. Special surface residues. The polypeptide chains that compose the threefold protrusion are shown (*19*). Residues that line the inside of the cylindrical structure about the fivefold axis are T152, S154, S156, T158, P160, P161, T162, V164, V166, N167, and L169.

Name	Position	Start	At top of protrusion	End
Loop 1	βB–βC	V84	V92, N93	D99
Loop 2*	βE–βF	R216	H222–T228	G235
Loop 3	βG–βH	P295	A300–F303	I306
Loop 4†	βG–βH	Y409	N421–N428, T433–N443	Y444

*Largest viral radius. †Greatest number of residues.

ing the best fit to the density. Thus, although most of the nucleic acid (as in other spherical viruses) is not seen, in CPV some DNA structure is visible where icosahedral symmetry has been imposed on it by internal binding to the coat protein. Icosahedrally ordered nucleic acid could occur when the internal surface of the icosahedral protein capsid contains a specific binding site for oligonucleotide sequences. The only other icosahedral virus structure for which ordered nucleic acid has been observed is that for the plant bean pod mottle virus (BPMV) (*40*). Both BPMV and CPV, in contrast to other viruses whose crystalline structures have been determined, readily form large amounts of empty particles in vivo (*41*). Thus, the protein and nucleic acid probably form weak interactions.

A total of 11 nucleotides were built into this density. Since this occurs in each of the 60 icosahedral asymmetric units of the virus, there are a total of 660 visible bases, or 13 percent of all of the encapsidated DNA. The built DNA structure conforms with allowed conformational angles (*39*), van der Waals contacts, and observed electron density. It contains riboses in both $C_{2'}$-*endo* and $C_{3'}$-*endo* conformations. Each of the 60 symmetry equivalent oligonucleotides contains a loop nestled into a pocket on the internal surface of the protein shell. Four bases are stacked on one side of the loop in a manner reminiscent of other nucleic acid structures. Surprisingly, the sixth, seventh, eighth, and tenth bases are neighbors in the stack because the ninth base points in the opposite direction (Fig. 4A). Two metal ions chelated between the DNA phosphates help to stabilize the loop. One lies inside a ring of six nucleotides (NA4 to NA9) that are chelated by four of the phosphates and Asn^{180}. No base pairing is observed, but rather the bases point outward toward the internal pocket of VP-2. There are only a few protein-DNA hydrogen bonds.

Although the electron density represents the average of 60 oligonucleotides within the genome, some bases can clearly be recognized as being predominantly either a purine or pyrimidine. For instance, base no. 3 has excellent density and shows that purines predominantly occupy this position, whereas base no. 7 would suffer steric collision with residue Tyr^{244} if it were a purine. Furthermore, there is some indication for differentiation between A and G or between T and C. Presumably van der Waals contacts are sufficient to endow some specificity for DNA binding. The preferred sequence is XYAYCTYRRRX [where X is A, G, T, or C; Y is T or C; and R is A or G; (*42*)]. A systematic search of the negative-strand CPV

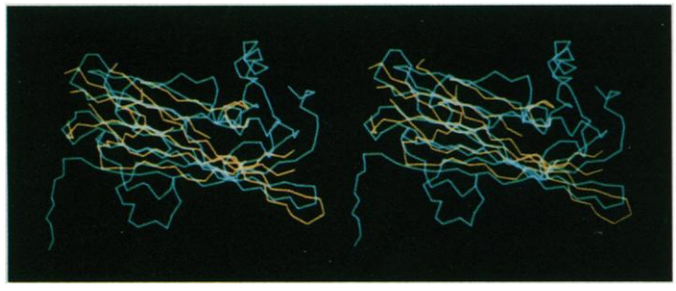

Fig. 5. Best superposition of the conserved β-barrel structure in CPV (yellow) and SBMV (blue). Although only the relevant polypeptide of CPV is shown, the whole of the SBMV-C subunit, apart from the disordered amino-terminal section, is shown.

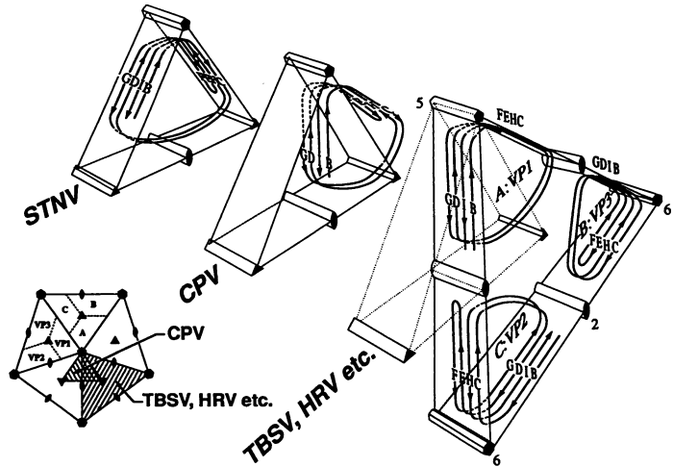

Fig. 6. Diagrammatic view of packing of subunits into an icosahedron. The positions of the four-stranded antiparallel sheets, BIDG and CHEF, are shown. Together they make the opposite sides of the β barrel. The letters in BIDG and CHEF represent the sequential names of the β strands within the polypeptide that make up the two sheets. (**Foreground**) Packing for $T = 3$ plant viruses (subunits A, C, and B are covalently identical) and pseudo $T = 3$ picorna- and comoviruses (VP-1, VP-2, and VP-3 are different proteins) within one icosahedral asymmetric unit. Icosahedral symmetry axes are shown. (**Middle**) One subunit of $T = 1$ CPV within its icosahedral asymmetric unit. The quasi–three- and twofold axes of the $T = 3$ viruses correspond to the icosahedral three- and twofold axes of CPV. The quaternary organization within the capsid is similar for the $T = 3$ viruses and for CPV, but not for satellite tobacco necrosis virus (STNV). (**Background**) One subunit of the $T = 1$ plant STNV. The quaternary organization differs to that of the $T = 3$ viruses, although the tertiary fold of the subunits remains similar.

nucleotide sequence revealed similar sequences but at a frequency no greater than would be expected by chance. Nevertheless, the DNA binding site may be important for packaging of viral DNA by the capsid protein.

Sequence and structure comparisons. The structure of the CPV coat protein was compared quantitatively (*43*) with that of VP1, VP2, and VP3 of human rhinovirus 14 (HRV14), subunit A of SBMV (Fig. 5), and satellite tobacco necrosis virus (STNV). STNV is the only known high-resolution $T = 1$ virus structure (*44*). Although the tertiary structure of the STNV coat protein fold is similar to CPV and other viruses (Table 3), the organization of the quaternary structure is quite different (Fig. 6). HRV14 represents a typical small RNA animal virus in both tertiary and quaternary structures and is similar to $T = 3$ plant viruses like tomato bushy stunt virus (*45*) and SBMV (*38*), to insect viruses such as black beetle virus (*46*), and to comoviruses (*47*). The in vitro assembly of SBMV subunits as a $T = 1$ empty shell shows similar subunit association as for the full $T = 3$ SBMV capsid (*48*) in the sense that the quasi-three- and twofold axes become icosahedral three-, and twofold axes and that the quasi-sixfold axes become icosahedral fivefold axes. Thus, the $T = 1$ CPV quaternary structure has considerable similarity to the $T = 1$ SBMV structure and hence also to the other $T = 3$ viruses (Fig. 6).

Comparisons of a few typical virus folds with CPV are summarized in Table 3. It is remarkable how faithfully the total topology of the eight-stranded β barrel in CPV superimposes on other viral folds. The aligned sequences shown in Fig. 7 are those derived from the "HOMOlogy" program (*29*). It is seen that the large insertion between βG and βH in CPV, which forms much of the threefold spikes, follows (in part) the FMDV loop insertion in VP-1 of HRV14. The "puff" insertion between βE and αB in VP-2 of HRV14 is also partly present in CPV (not shown in Fig. 7) and forms another component of the threefold protrusion in CPV. There is a sizable cavity in CPV corresponding to the drug-binding pocket in HRV-VP1 (*49*), although (unlike in HRV14) it is not accessible from the virus exterior. The total number of amino acids in VP-2 of CPV that can be made topologically equivalent with other virus folds is in the order of ~100 residues. This number represents ~20 percent of the amino acids in the β-barrel domain of CPV but as much as 62 percent of HRV14-VP2. Similar comparisons have been made among structures of other protein folds (*43, 50*). If these are used as benchmarks, it is probable that parvoviruses, picornaviruses, plant RNA $T = 3$ viruses, and insect RNA $T = 3$ viruses are likely to have diverged from a common ancestral structure. A similar but more distant relation exists between the tertiary structure of other viral capsid proteins such as the double-stranded DNA adenovirus hexon capsid protein (*51*). The only exception to the use of the eight-stranded β-barrel motif in the construction of icosahedral viruses is the bacteriophage MS2 (*52*). With the limited information available at this time, it might be tentatively concluded that those isometric RNA and DNA viruses that can infect eukaryotic cells may have evolved from a common viral precursor.

The position of secondary structural elements of CPV in the antiparallel β-barrel motif is shown in Fig. 7. Residues that line the internal hydrophobic pocket of the β barrel show greater conservation. Residue Gly^{496} in βH of CPV, corresponding to Gly^{1222} in HRV14-VP1 and Gly^{3185} in HRV14-VP3, is conserved in all known VP3 picornavirus sequences (*53*). The high degree of conservation of these residues, as well as some of the hydrophobic residues within the β barrel such as CPV Leu^{173} and Leu^{498}, may indicate an essential folding requirement to produce the β-barrel structure.

Cylindrical structure about the fivefold axes. The cylindrical structure around the fivefold axes (Fig. 2) consists of an antiparallel β ribbon running from Thr^{152} to Pro^{160} and back again from Pro^{161} to Leu^{169}. The five antiparallel β ribbons, generated by each fivefold axis, are separated sufficiently far to rule out hydrogen bonding between them. The smallest inside diameter of the β cylinder is 8 Å (at a viral radius of 104 Å, corresponding to residue Leu^{169}). Many of the residues lining the inside of the cylinder (Table 4) are small and hydrophilic. In full CPV capsids there is electron density along

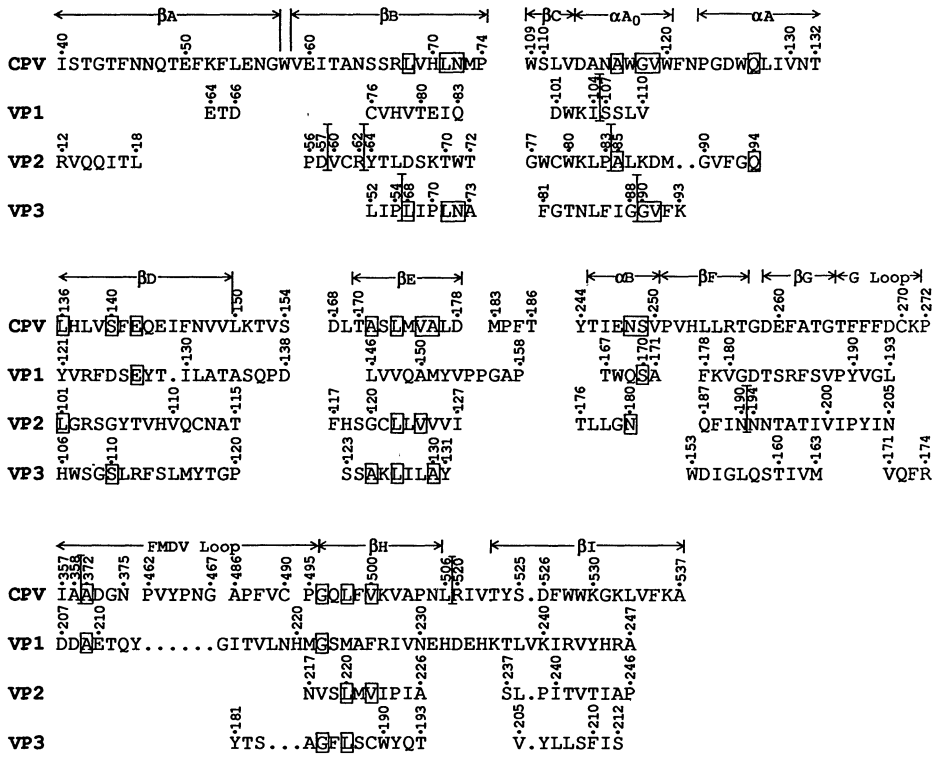

Fig. 7. Amino acids and secondary structural elements in the antiparallel β-barrel domain of CPV that can be structurally superimposed on VP1, VP2, or VP3 of HRV14 according to the technique of Rao and Rossmann (*29*). Boxed residues are conserved in at least one of the aforementioned sequences. The "FMDV loop" refers to a structural element found in VP1 of HRV14 (*31*).

the fivefold axes from a viral radius of 90 Å to 107 Å that is connected to residue 38. Much of this density is as high as the other polypeptide density. Although density along any of the symmetry axes is prone to error (due to symmetry averaging), nevertheless, it would seem probable that some of the amino ends of VP-2 or VP-3 are folded through the cylindrical structure. Presumably, if this is the case, then this fold is induced onto some of the viral proteins during the assembly of the particle, rather than the polypeptide being threaded through the cylinder after viral assembly is completed. In general, only one out of the five of the amino ends of VP-2 or VP-3 would be able to pack into the cylindrical structure. Support for this interpretation of the electron density map comes from the absence of any density along the fivefold axes in the map of empty CPV particles, which is consistent with the observation that only VP-2's in full particles can be cleaved by trypsin (*12, 13*). Apparently the presence of the DNA causes some of the amino ends of VP-2 or VP-3 to be outside of the particle.

The conserved polyglycine sequence immediately prior to the clearly visible Val^{38} residue would be required both for flexibility and smallness in order to pass through the β cylinder. Cleavage at Arg^{16} or Arg^{19} would expose a group of moderately hydrophobic glycines. This process may be required for translocation of the virus across the plasma membrane (*9, 12*). An analogous situation exists in picornaviruses that lose the small internal VP-4 protein during the initial stages of infection. This protein is myristoylated in at least some picornaviruses (*54*) and may be required for membrane binding. The hydrophobic myristoylate is situated near the fivefold axes. Thus, the initial step of losing the VP-4 may be along the fivefold vertices (*54*). As the NS-1 protein has been shown to be external to the virion but attached to the 5′ end of DNA (*15*), it is reasonable to assume that the DNA might also pass through the cylindrical structure, although the phosphates, sugars, and bases would have a larger mean radius than the polyglycine chain. Again, a comparison with picornaviruses is relevant, as their RNA is released from the capsid after the departure of the VP-4 (*55*).

Antigenic site, host range differentiation, and hemagglutination. There are only a few results detailing the amino acids associated with binding of neutralizing antibodies to CPV or homologous viruses. Residue Ala^{300} of CPV defines a common epitope in CPV and FPV for monoclonal antibodies raised to CPV (*6*). This residue is on the extreme outside surface of the virus and is at the tip of one of the loops forming the threefold spikes. Another residue associated with a CPV epitope is Asn^{93} (*6*), which is also on the extremity of the spike on the threefold axes (Table 4). Thus, as in picornaviruses, the neutralizing epitopes appear to be on the most exposed surfaces of the virions. Residue Asn^{93} has also been associated with host range differentiation between CPV and FPV (*6*), suggesting that the differentiation might be associated with the antigenic properties of the viruses.

Recombination mapping has indicated that residues Asn^{375}, Arg^{377}, Val^{562}, Ser^{564}, and Gly^{568} are associated with the ability of CPV to hemagglutinate or alter pH dependence of hemagglutination (*56*). All but residue 564 are buried (Fig. 1C) and are therefore unlikely to effect hemagglutination other than they might indirectly cause surface conformational changes.

The canyon. There are two major depressions (Fig. 1, B and C) on the viral surface; one is on the twofold axes, and the second encircles each of the fivefold cylindrical structures and is reminiscent of the canyon found in rhino- (*31*) and other picornaviruses. A variety of data indicates that the canyon is the site of receptor attachment in rhinoviruses (*57*). The residues lining the canyon surface are more conserved than those elsewhere on the surface in rhino-, entero-, and cardioviruses (*58*). The depression should hinder the binding of antibodies to the canyon but would allow virus binding to the receptor. Hence, the virus can conserve its ability to bind to a specific receptor while still rapidly changing its more exposed surfaces to avoid host immune surveillance. Of the two depressions on CPV, the canyon would seem to be the more likely site of receptor attachment both because it is narrow and, therefore, a better deterrent to antibody and perhaps because of its similarity in position to the rhinovirus canyon.

REFERENCES AND NOTES

1. G. Kurtzman *et al.*, *N. Engl. J. Med.* **321**, 519 (1989).
2. L. J. Anderson and T. J. Torok, *ibid.*, p. 536.
3. M. J. Studdert, in *Handbook of Parvoviruses*, P. Tijssen, Ed. (CRC Press, Boca Raton, FL, 1990), vol. 2, pp. 27–32.
4. C. R. Parrish *et al.*, *J. Gen. Virol.* **69**, 1111 (1988).
5. C. R. Parrish, P. H. O'Connell, J. F. Evermann, L. E. Carmichael, *Science* **230**, 1046 (1985).
6. C. R. Parrish, C. F. Aquadro, L. E. Carmichael, *Virology* **166**, 293 (1988).
7. S. Basak and R. W. Compans, *J. Virol.* **63**, 3164 (1989).
8. P. Tattersall and S. F. Cotmore, in *Parvoviruses and Human Disease*, J. R. Pattison, Ed. (CRC Press, Boca Raton, FL, 1988), pp. 5–41.
9. P. Tattersall and E. M. Gardiner, in *Handbook of Parvoviruses*, P. Tijssen, Ed. (CRC Press, Boca Raton, FL, 1990), vol. 1, pp. 111–122; P. Tattersall and S. F. Cotmore, in *ibid.*, pp. 123–140.
10. C. V. Jongeneel, R. Sahli, G. K. McMaster, B. Hirt, *J. Virol.* **59**, 564 (1986).
11. D. L. C. Caspar and A. Klug, *Cold Spring Harbor Symp. Quant. Biol.* **27**, 1 (1962).
12. G. M. Clinton and M. Hayashi, *Virology* **74**, 57 (1976).
13. P. Tattersall, A. J. Shatkin, D. C. Ward, *J. Mol. Biol.* **111**, 375 (1977).
14. D. C. Kelly and R. M. Elliot, *J. Virol.* **21**, 408 (1977).
15. S. F. Cotmore and P. Tattersall, *ibid.* **63**, 3902 (1989).
16. A. I. Ranz, J. J. Manclús, E. Díaz-Aroca, J. I. Casal, *J. Gen. Virol.* **70**, 2541 (1989).
17. S. L. Rhode III, *J. Virol.* **54**, 630 (1985).
18. A. P. Reed, E. V. Jones, T. J. Miller, *ibid.* **62**, 266 (1988).
19. Abbreviations for the amino acid residues are A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.
20. C. R. Wobbe, S. Mitra, V. Ramakrishnan, *Biochemistry* **23**, 6565 (1984).
21. M. Luo, J. Tsao, M. G. Rossmann, S. Basak, R. W. Compans, *J. Mol. Biol.* **200**, 209 (1988).
22. B. J. Carter and P. J. Tattersall, in *Animal Virus Structure*, M. V. Nermut and A. C. Steven, Eds. (Elsevier, Amsterdam, 1987), pp. 325–334.
23. CPV type 2 (the first strain to become established in dogs in 1978) was propagated in and purified from the canine A72 cell line by a modified procedure (*21*) of P. R. Paradiso [*J. Virol.* **39**, 800 (1981)]. The procedure was further modified by suspending the precipitated virus in 10 mM tris-HCl buffer at pH 7.5 containing 0.1 percent sarcosine and 1 mM EDTA and then pelleting the virus through a 30 percent sucrose cushion. The pellet was then resuspended and centrifuged in a CsCl gradient as before. In general, there were two bands with an A_{260}/A_{280} ratio (absorbance at 260 and 280 nm) of 1.40 for the heavier and 0.70 for the lighter band. Electron microscopy confirmed that these bands corresponded to particles that were full (110*S*) or empty (70*S*) of DNA, respectively. Only the full particles were infectious. Both types of particles could be crystallized. In addition to the previously reported monoclinic crystals that were favored at higher virus concentration (15 mg/ml), at least three other crystal forms were frequently observed and characterized. Seeding was occasionally successful in increasing the yield of monoclinic crystals. The full and empty particles gave isomorphous crystals in all of the observed crystal forms.
24. S. Kim, *J. Appl. Crystallogr.* **22**, 53 (1989).
25. M. S. Chapman, unpublished results.
26. M. G. Rossmann, *J. Appl. Crystallogr.* **12**, 225 (1979).
27. M. G. Rossmann, A. G. W. Leslie, S. S. Abdel-Meguid, T. Tsukihara, *ibid.*, p. 570.
28. M. G. Rossmann and D. M. Blow, *Acta Crystallogr.* **15**, 24 (1962).
29. S. T. Rao and M. G. Rossmann, *J. Mol. Biol.* **76**, 241 (1973).
30. M. G. Rossmann, *Acta Crystallogr.* **A46**, 73 (1990); W. P. J. Gaykema *et al.*, *Nature* **309**, 23 (1984).
31. M. G. Rossmann *et al.*, *Nature* **317**, 145 (1985).
32. The final correlation coefficients C for 15 to 3.25 Å resolution data: 15.0 to 8.63 Å, 0.947; 8.63 to 6.68 Å, 0.940; 6.68 to 5.64 Å, 0.921; 5.64 to 4.98 Å, 0.896; 4.98 to 4.50 Å, 0.864; 4.50 to 4.14 Å, 0.817; 4.14 to 3.85 Å, 0.751; 3.85 to 3.62 Å, 0.706; 3.62 to 3.42 Å, 0.616; and 3.42 to 3.25 Å, 0.379; where

$$C = \frac{\Sigma(\langle F_o\rangle - F_o)(\langle F_c\rangle - F_c)}{\{\Sigma(\langle F_o\rangle - F_o)^2 \cdot \Sigma(\langle F_c\rangle - F_c)^2\}^{1/2}}$$

and where $\langle F_o\rangle$ and $\langle F_c\rangle$ are the mean observed and calculated amplitudes in each shell. The F_c values were obtained by Fourier backtransformation of the averaged electron density.
33. T. A. Jones, *J. Appl. Crystallogr.* **11**, 268 (1978); _____ and S. Thirup, *EMBO J.* **5**, 819 (1986).
34. The amino acid numbering sequence refers to VP-2 of CPV (*6, 18*).
35. T. L. Blundell and L. N. Johnson, in *Protein Crystallography* (Academic Press, New York, 1976), p. 224.

36. M. G. Rossmann and J. E. Johnson, *Annu. Rev. Biochem.* **58**, 533 (1989).
37. Foot and Mouth Disease Virus (FMDV); R. Acharya *et al.*, *Nature* **337**, 709 (1989).
38. Southern Bean Mosaic Virus (SBMV); C. Abad-Zapatero *et al.*, *ibid.* **286**, 33 (1980).
39. W. Sanger, *Principles of Nucleic Acid Structure* (Springer-Verlag, New York, 1984); S. Arnott, D. W. L. Hukins, S. D. Dover, W. Fuller, A. R. Hodgson, *J. Mol. Biol.* **81**, 107 (1973).
40. Z. Chen *et al.*, *Science* **245**, 154 (1989).
41. We are grateful to J. E. Johnson for pointing this out to us.
42. IUB Nomenclature Committee, *Eur. J. Biochem.* **150**, 1 (1985).
43. B. W. Matthews and M. G. Rossmann, *Methods Enzymol.* **115**, 397 (1985).
44. L. Liljas *et al.*, *J. Mol. Biol.* **159**, 93 (1982); T. A. Jones and L. Liljas, *ibid.* **177**, 735 (1984).
45. S. C. Harrison, A. J. Olson, C. E. Schutt, F. K. Winkler, G. Bricogne, *Nature* **276**, 368 (1978).
46. M. V. Hosur *et al.*, *Proteins* **2**, 167 (1987).
47. C. V. Stauffacher *et al.*, in *Crystallography in Molecular Biology*, D. Moras, J. Drenth, B. Strandberg, D. Suck, K. Wilson, Eds. (Plenum, New York, 1987), pp. 293–308.
48. J. W. Erickson, A. M. Silva, M. R. N. Murthy, I. Fita, M. G. Rossmann, *Science* **229**, 625 (1985).
49. T. J. Smith *et al.*, *ibid.* **233**, 1286 (1986).
50. M. G. Rossmann, *BioEssays* **7**, 99 (1987).
51. M. M. Roberts, J. L. White, M. G. Grütter, R. M. Burnett, *Science* **232**, 1148 (1986).
52. K. Valegård, L. Liljas, K. Fridborg, T. Unge, *Nature* **345**, 36 (1990).
53. A. C. Palmenberg, in *Molecular Aspects of Picornavirus Infection and Detection*, B. L. Semler and E. Ehrenfeld, Eds. (American Society for Microbiology, Washington, DC, 1989), pp. 211–241.
54. M. Chow *et al.*, *Nature* **327**, 482 (1987).
55. F. Koch and G. Koch, *The Molecular Biology of Poliovirus* (Springer-Verlag, New York, 1985).
56. C. R. Parrish, G. Burtonboy, L. E. Carmichael, *Virology* **163**, 230 (1988).
57. M. G. Rossmann, *J. Biol. Chem.* **264**, 14587 (1989); R. J. Colonno *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 5449 (1988).
58. M. G. Rossmann and A. C. Palmenberg, *Virology* **164**, 373 (1988).
59. We thank C. Bugg for allowing one of us (J.T.) to use graphic facilities at the University of Alabama and M. Carson for help in its use; G. Kamer and S. Kim for providing help and instruction with programming problems; M. R. N. Murthy for a month of active and very helpful participation in the processing of CPV data; S. Basak for samples of CPV inoculant; S. Kim and R. McKenna for helpful discussions about film processing and the latter also for help with low-resolution area detector data collection; a large group of helpers on numerous trips to collect data at synchrotrons (R. McKenna, A. Prongay, S. Krishnaswamy, J. Bibler, L. Tong, P. Willingmann, and D. Xia); the staff of the CHESS and Daresbury synchrotrons for their dedicated help; J.-Y. Sgro and I. Minor for help in the preparation of Figs. 1B and 1C, respectively; and, finally, H. Prongay and S. Wilder for help in preparation of the manuscript. The work was supported by grants to M.G.R. from the National Institutes of Health and the National Science Foundation and by a Markey Foundation grant supporting the expansion of structural studies at Purdue University.

5 November 1990; accepted 17 January 1991