Acta Cryst. (1993). D**49**, 572–579

Determination and Refinement of the Canine Parvovirus Empty-Capsid Structure

BY HAO WU,* WALTER KELLER† AND MICHAEL G. ROSSMANN‡

Department of Biological Sciences, Purdue University, West Lafayette, IN 47907, *USA*

(*Received* 9 *February* 1993; *accepted* 30 *June* 1993)

Abstract

The canine parvovirus (CPV) empty-capsid structure has been determined and refined to 3.0 Å resolution in the tetragonal space group $P4_32_12$ with cell dimensions $a = b = 254.5$ and $c = 795.0$ Å. The successful structure determination shows that reasonably good diffraction data were obtained in spite of the very long c axis. The structure was solved by molecular replacement using the electron density of CPV full particles in a monoclinic space group. The phases were refined by non-crystallographic symmetry averaging. The structure refinement was carried out by using the programs *PROLSQ* and *X-PLOR*. The final R factor for the structure that included 85 water molecules per icosahedral asymmetric unit was 21.1% for reflections between 6.0 and 3.0 Å resolution with an r.m.s. deviation of bond lengths of 0.020 Å from ideal values. The structure of CPV empty capsids showed conformational differences with respect to full capsids at a region where icosahedrally ordered DNA in full particles interacts with the capsid protein. It also confirmed the absence of density along the fivefold axis in the CPV empty-particle structure in contrast to the situation in CPV full particles.

Introduction

Parvoviruses are a group of small icosahedral single-stranded DNA viruses that infect many animals including humans (Berns, 1984). Worldwide canine parvovirus (CPV) infection first occurred around 1978 (Parrish, Have, Foreyt, Evermann, Senda & Carmichael, 1988) and is now endemic in most populations of wild and domestic dogs (Parrish, Aquadro, Strassheim, Evermann, Sgro & Mohammed, 1991). CPV capsids consist of 60 copies of a combination of VP1 (82.3 kDa), VP2 (67.3 kDa) and VP3 (63.5 kDa) (Paradiso, Rhode & Singer, 1982; Cotmore & Tattersall, 1987). The larger proteins contain the entire sequence of the smaller proteins with unique amino-terminal extensions. VP1 and VP2 are primary translational products while VP3 is the proteolytic product of VP2 (Tattersall & Cotmore, 1988). CPV full particles (DNA-containing) have all three viral coat proteins while CPV empty capsids (non-DNA-containing) have only VP1 and VP2. In addition, trypsin treatment of CPV full particles cleaves VP2 to a VP3-like peptide, whereas the same treatment of CPV empty particles does not convert VP2 to VP3 (Clinton & Hayashi, 1976; Tattersall, Shatkin & Ward, 1977). Thus, the tryptic sites of VP2 are apparently exposed in CPV full particles but not in CPV empty particles.

The propagation and purification of the CPV-2 serotype was described previously (Paradiso, 1981; Luo, Tsao, Rossmann, Basak & Compans, 1988; Tsao *et al.*, 1991). A CsCl density gradient was used to separate the full (density 1.41–1.45 g cm^{-3}) and empty (density 1.32 g cm^{-3}) particles. Both CPV full and empty particles were crystallized using the hanging-drop vapor-diffusion method in Tris–HCl buffer at pH 7.5 containing 0.75% PEG8000 and 8 mM $CaCl_2$ (Luo *et al.*, 1988; Tsao *et al.*, 1991). Different crystal forms grew from the same crystallization conditions. Sufficient crystals were obtained for the structure determination of CPV full particles in a monoclinic space group and of CPV empty particles in a tetragonal space group. The structure solution to CPV full particles has been reported (Tsao *et al.*, 1992). In this paper we describe the structure determination and refinement of CPV empty capsids in the tetragonal space group $P4_32_12$ ($a = b = 254.5$ and $c = 795.0$ Å).

Data collection

Diffraction data were collected using oscillation photography at the A1 beamline of the Cornell High Energy Synchrotron Source (CHESS) with a wavelength of about 1.56 Å. Tetragonal CPV empty-particle crystals diffracted up to 2.5 Å resolution. The CPV tetragonal unit cell has a long c axis of about 800 Å (see below). If this axis was in the plane

* Current address: Department of Biochemistry and Molecular Biophysics, College of Physicians and Surgeons of Columbia University, 630 West 168th Street, New York, New York 10032, USA.

† Current address: Institut für Molekularbiologie und Biophysik, ETH-Zürich/Hönggerberg, CH-8092 Zürich, Switzerland.

‡ To whom correspondence should be addressed.

Table 1. *Summary of the diffraction data*

(*a*) Data processed

	No. of films	No. of observations	No. of reflections	$F^2/\sigma(F^2)$	R_{merge} (%)*
Empty capsids	43	391×10^3	243×10^3	2.5	11.7 ('old')† 10.5 ('new')†
Full capsids	48	174×10^3	135×10^3	2.5	15.6

(*b*) Percentage of theoretically possible data in each resolution shell

Resolution (Å)	∞–30	30–15	15–10	10–7.5	7.5–5	5–3.5	3.5–3	3–2.75
Empty capsids (%)	59	72	70	68	65	49	26	11
Full capsids (%)	44	58	53	48	40	26	14	2

* $R_{merge} = (\sum_h \sum_i |\langle F_h^2 \rangle - F_{hi}^2| / \sum_h \sum_i \langle F_h^2 \rangle) \times 100$ where there are i observations F_{hi}^2 of reflection h resulting in the mean value of $\langle F_h^2 \rangle$.
† 'Old' refers to the initial data processing with slightly different cell dimensions; 'new' refers to the final data processing with unified cell dimensions.

of the film during an initial exposure, extremely closely spaced spots could be seen in Polaroid pictures. To collect data from crystals in this orientation (a '*c**' film), the crystal-to-film distance was increased from 100 (for an '*a***' film) to 150 mm in order to separate the reflection maxima, although this reduced the extreme recording limit of resolution from 2.5 to 3.2 Å. An oscillation angle of 0.4° was used for both the '*a***' and '*c**' films. In retrospect, a smaller oscillation angle should have been used for the '*a***' films to reduce the severe lunar overlap due to the long *c* axis. Some data were also collected for isomorphous tetragonal CPV full-particle crystals.

The Kodak films (120 × 120 mm) used to record the diffraction pattern have double emulsion layers coated on each side of the film, causing overlap of the closely spaced spots along the *c** direction beyond about 5 Å resolution. This is a result of the displacement of the record on either side of the film due to the inclination of the diffracted rays to the film normal.

Data processing

The films were indexed with the auto-indexing procedure of Kim (1989) and processed and post-refined using the Purdue film-processing package (Rossmann, 1979; Rossmann, Leslie, Abdel-Meguid & Tsukihara, 1979). Because of the severe lunar interpenetration in '*a***' films, about one-third of the total reflections were partially overlapped. However, many of these reflections could be deconvoluted using the previously learned profile.† The overall *R* factor was 11.7% for reflections with $F^2 > 2.5\sigma(F^2)$. The data completeness was roughly 50–60% at 3.5 Å resolution and decreased to about 20% at 3.0 Å resolution. Inside 3.5 Å resolution about half of the reflections had more than one observation, while between 3.5 and 3.0 Å resolution only a quarter of the reflections had multiple observations. Data were also collected for full particles (Table 1).

There is a systematic lack of data along the *c** axis at high resolution because the '*c**' films were taken at a longer crystal-to-film distance (150 mm). This uneven distribution of reflections in reciprocal space did not seem to be a problem in the structure determination, presumably because of the presence of high non-crystallographic symmetry.

The '*a***' films gave a roughly orthogonal unit cell in the auto-indexing procedure, with *a* and *b* close to 255 and *c* around 801 Å. The '*c**' films gave similar *a* and *b* axial lengths, but somewhat different *c* axis lengths of around 795 Å. Attempts to process the two sets of films with the averaged cell dimensions gave worse processing statistics and, hence, the two types of films were initially processed using slightly different cell dimensions. Initial post-refinement of cell parameters was not able to obtain consistent cell dimensions. The subsequent structure determination, unless otherwise stated, used the data set created from films processed with two sets of slightly different cell dimensions. Cell dimensions of $a = b = 255$ and $c = 801$ Å were assumed in the structure determination by molecular replacement averaging. This data set had sufficient quality to enable the successful determination of the structure (see below).

After the structure was solved, a second attempt was made to search for consistent and more accurate cell dimensions. It was found that the cell dimensions ($a = b = 255.5$ and $c = 797.5$ Å) representing the compromise between the two sets of films gave worse processing statistics but were accurate enough to permit the post-refinement on cell parameters to converge, giving refined cell dimensions of $a = b = 254.5$ and $c = 795.0$ Å. Both processing statistics and scaling *R* factors were improved significantly with the new cell dimensions. A new data set was produced with an *R* factor of 10.5% (Table 1). The larger error found in the length of the *c* axis using the '*a***' films may be due to the close alignment of the *c* axis with the X-ray beam (within 25°), thus

† A reflection was accepted if the sum of the optical densities from neighboring overlapped reflections was less than 30% of its total intensity. The optical densities contributed by the neighbors were based on a prior approximate estimate of the intensities and a knowledge of the mean reflection profile.

correlating the c axial length with crystal-to-film distance, and due to the limited sampling of reflections along the c^* direction for auto-indexing.

Resolution-dependent film scaling

Individual films from a single crystal or from different crystals would be expected to each have different behavior as a function of resolution due to crystal decay and differing absorption effects. This implies that the use of resolution-dependent scale factors in film scaling might yield better data than the use of a single scale factor. Although the available program (Rossmann *et al.*, 1979) does permit the refinement of radially dependent film-to-film scale factors, the function used does not relate well to the observations. Both theoretical and empirical methods have been used to approach the problem (Huber & Kopfmann, 1969; Schwager, Bartels & Huber, 1973; Matthews & Czerwinski, 1975; Stuart & Walker, 1979). Here, scale factors were calculated in resolution shells and then were correlated by the use of a smoothing function.

The scale factor that divides $F_{h,i}^2$ for the ith resolution shell of a given film was defined as

$$g_i^{\text{obs}} = \sum_h \bar{F}_h^2 F_{hi}^2 / \sum_h (\bar{F}_h^2)^2,$$

where $\bar{F}_h^2$ is the intensity of reflection h from the data set created using single scale factors and F_{hi}^2 represents an observation on the given film within the ith resolution shell. The whole resolution range was divided into shells of equal reciprocal volume. Any resolution shell that had too few reflections was merged with neighboring resolution shells until the total number of reflections was more than a selected cutoff value (*e.g.* 100).

The variation of scale factors in successive resolution shells was smoothed using either a polynomial ($g_i^{\text{calc}} = a + bx + cx^2 + dx^3 + ex^4$ where x = 1/resolution) or a Gaussian [$g_i^{\text{calc}} = k\exp(-Bx^2)$] function (Fig. 1). The R factor from local scaling was 11.7 and 11.8% for polynomial and Gaussian function fitting, respectively, whereas the R factor from single scale factor scaling was 12.2% (Table 2*a*). The improvement in R factor was most significant for films which had unusually large intensity falloffs and were often the last few exposures of a crystal (Table 2*b*).

Space group, particle orientation and particle position

The cell dimensions from auto-indexing suggested that the crystals might have tetragonal symmetry. This was confirmed by the intensity agreement of 4/*mmm*-related reflections in reciprocal space. Systematic absences showed that the lattice was primitive and suggested the presence of a 4_1 or 4_3 axis along c and a 2_1 axis along a and b. Hence, the space group was probably either $P4_12_12$ or its enantiomorph $P4_32_12$. A calculation of V_M (Matthews, 1968) showed that there could only be four particles in the unit cell ($V_M = 2.2$ Å^3 Da^{-1}), which indicated that the virus particle has to sit on a crystallographic twofold axis along an *ab* face diagonal, giving rise to 30-fold non-crystallographic redundancy.

The orientation of the virus particle in the unit cell relative to a 'standard' orientation (Fig. 2) was determined by general (Rossmann & Blow, 1962) and locked (Rossmann, Ford, Watson & Banaszak, 1972; Tong & Rossmann, 1990) rotation functions. The locked rotation function was calculated using data betwen 6 and 5 Å resolution by aligning a twofold axis of the 'standard' icosahedron with a crystallographic twofold axis and rotating about the common twofold axis (Eulerian angles $\theta_1 = 45$, $\theta_3 = 0°$) (Fig. 2). The highest peak was found to be at $\theta_2 = 12.94°$ (Fig. 2), corresponding to a rotation matrix of

$$[E] = \begin{pmatrix} 0.7071 & 0.7071 & 0.0000 \\ -0.6891 & 0.6891 & 0.2241 \\ 0.1585 & -0.1585 & 0.9746 \end{pmatrix}.$$

This matrix rotates the orientation of one of the particles in the crystal unit cell into the 'standard' icosahedron (Fig. 2).

By considering the packing of spheres into the crystal unit cell, a virus particle position near (0.5, 0.5, 0), or equivalently near (0.0, 0.0, 0), was favored. Using the 3.6 Å resolution electron-density map of

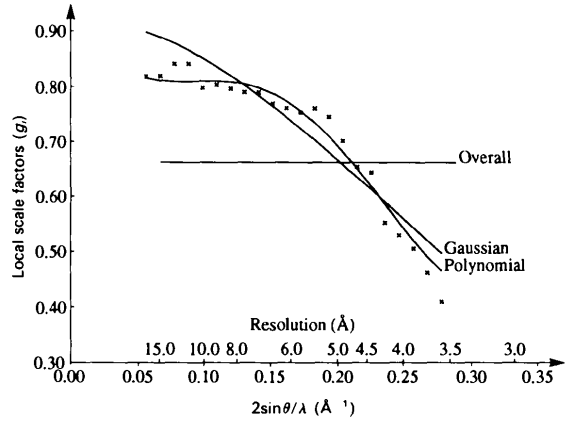

Fig. 1. The local scale factors, g_i, and their fit to a polynomial and a Gaussian function for one selected film. The local scale factor was determined by scaling to the complete data set produced by employing a single ('overall') scale factor per film. If F_h^2 is the intensity of reflection h from the data set created by using only a single scale factor for each film, and F_{hi}^2 is the observation on film i then $g_i^{\text{obs}} = [\sum_h \bar{F}_h^2 F_{hi}]/[\sum_h (\bar{F}_h^2)^2]$. The g_i^{obs} values are shown as crosses.

Table 2. *Local scaling*

(*a*) Improvement in R_{merge}(%)* with local scale factors

Resolution (Å)	∞–15	15–7.5	7.5–5	5–3.8	3.8–3	Overall
Single scale factor scaling	7.2	7.3	8.7	13.5	19.9	12.2
Polynomial scaling	5.6	5.8	8.1	13.0	19.2	11.7
Gaussian scaling	5.6	5.9	8.3	13.1	19.3	11.8
No. of reflections	2162	20613	53154	58757	26313	160999

(*b*) Variation of Gaussian scale-factor† parameter B with increased exposure of crystal

Successive film exposures	B
1	−0.154
2	−0.010
3	+0.267
4	+0.622

*See Table 1 for definition.

† The Gaussian scale factor was defined as $k\exp(-Bx^2)$ where x = 1/resolution.

the CPV full-particle structure, translation-function searches (Argos & Rossmann, 1980) were performed with data between 8.0 and 7.0 Å resolution for both the $P4_32_12$ and $P4_12_12$ space groups. The searches showed that the particle position was at (0.533, 0.533, 0) and that the correct space group was $P4_32_12$ as judged by the significantly different heights of the peaks (18.1 *versus* 10.3σ) in the two space groups. When the searches were repeated at a higher resolution (between 4.2 and 4.0 Å), the differentiation between the two space groups was even greater (Fig. 3), with peak heights of 14.2 *versus* 5.2σ. The resultant packing organization shows that there is a tendency for the particles to pack with their fivefold vertices in contact (Fig. 4).

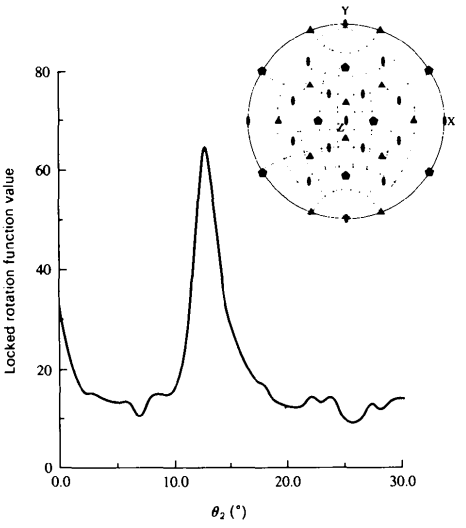

Fig. 2. The locked rotation function of tetragonal CPV. The standard icosahedral orientation is given in the top right-hand corner insert. The standard icosahedron is rotated to place its x axis along the crystallographic (110) direction and rotated again about the common superimposed axis by setting the Eulerian angles $\theta_1 = 45$, $\theta_3 = 0°$ and allowing θ_2 to change. The locked rotation function value is the sum of the general rotation function for each of the icosahedral symmetry axes.

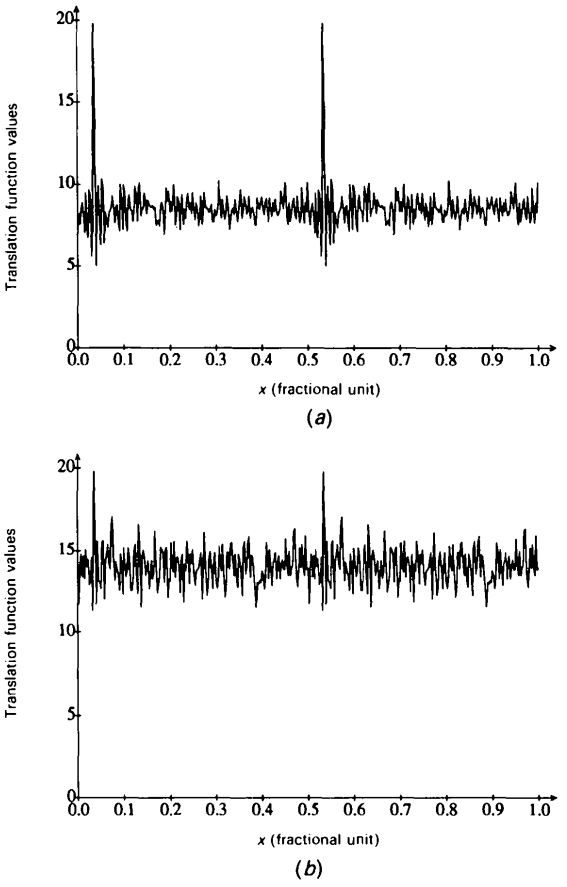

Fig. 3. Translation function searches at 4.2 to 4.0 Å resolution. (*a*) In $P4_32_12$ space group. (*b*) In $P4_12_12$ space group. Note that the positions at $x = 0.03$ and 0.53 are symmetry related and correspond to an origin shift of (1/2, 1/2, 0). The translation function was defined as $\sum_h (F_h^{obs})^2 (F_h^{calc})^2$, where F_h^{calc} was derived from the electron density of CPV full particles in a monoclinic space group placed into the tetragonal unit cell corresponding to the orientation found from the rotation function.

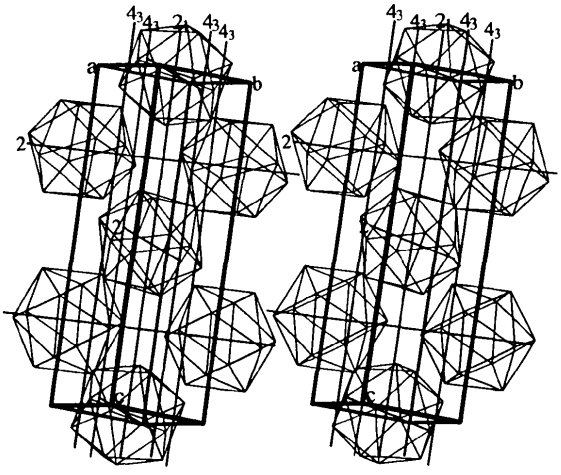

Fig. 4. Packing of icosahedral particles in the tetrahedral cell.

Molecular replacement averaging

Initial phases, between 15 and 6 Å resolution, obtained by orienting and positioning the CPV full-particle electron density correctly into the unit cell, were refined by averaging over the 30-fold non-crystallographic symmetry using the program *MolRep* (Johnson, 1978). A spherical shell with an inner radius of 75 Å and an outer radius of 140 Å was used to define the volume within which the density was averaged. The same radii had been used for the averaging of the CPV full-particle structure in the $P2_1$ space group (Tsao *et al.*, 1991). Tangential planes were applied at the middle of each inter-particle vector to limit the overlapping spheres. Densities outside the averaging mask were flattened to zero.

Phases calculated from the averaged map were combined with the observed structure-factor amplitudes. The agreement between the observed and calculated structure-factor amplitudes for each reflection, the molecular replacement figure of merit, was evaluated by the geometric mean of Sim (Sim, 1959, 1960) and Rayment (Rayment, 1983) weighting according to the formula

$$w = (w_{\text{Sim}} \times w_{\text{Rayment}})^{1/2}.$$

Calculated structure factors were assigned the average figure of merit of the shell. The molecular replacement figure of merit was used as a weighting term in the electron-density map calculation of the next averaging cycle. Unobserved reflections were approximated by calculated structure factors and were included in the next cycle of averaging, weighted by the mean molecular replacement figure of merit in the resolution shell. This is important because of the relative incompleteness of the observed data.

After averaging had converged at a given resolution, initial phases derived from the monoclinic CPV electon-density map, instead of those obtained by Fourier back-transformation of the current averaged map, were incorporated into the structure factors in the extended part of the resolution. The presence of approximate initial phases permitted the resolution extension from 6.0 to 3.5 Å in a few large steps with three or four cycles of refinement at each resolution step. The final overall correlation coefficient and R factor between observed and back-transformed structure factors were 0.84 and 20.8%, respectively (Fig. 5).

An averaged electron-density map which was just large enough to cover one icosahedral asymmetric unit was calculated using data between 15.0 and 3.5 Å resolution. The atomic model built for CPV full particles in the $P2_1$ space group was modified using *FRODO* (Jones, 1978) on a PS390 computer according to the electron density for CPV empty capsids in the $P4_32_12$ space group. Phases for reflections between 3.5 and 3.0 Å resolution were calculated from the modified atomic model and combined with the previous molecular replacement phases from 15.0 to 3.5 Å resolution. Cycles of averaging were then carried out at 3.0 Å resolution using the program *ENV* (Rossmann *et al.*, 1992) with a molecular mask generated by this program. After 12 cycles of averaging at 3.0 Å resolution, the correlation coefficients converged to somewhat higher values in all resolution shells that those obtained from averaging using merely a spherical mask (Fig. 6). The results showed the advantage of defining a molecular mask over an approximate spherical mask. A map

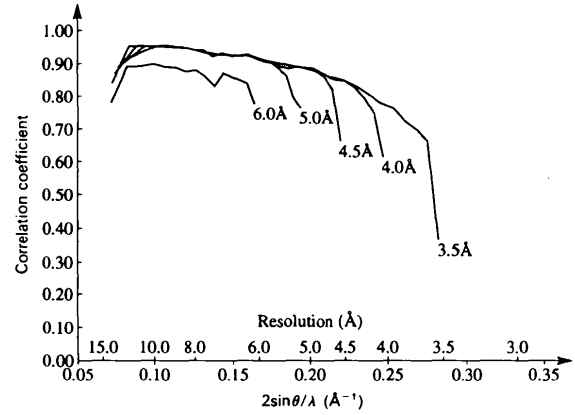

Fig. 5. Correlation coefficient as a function of resolution during the phase extension from 6 to 3.5 Å resolution. The correlation coefficient is defined as $CC = [\sum_h (F_h^{obs} - \langle F_h^{obs} \rangle)(F_h^{calc} - \langle F_h^{calc} \rangle)]/[\sum_h (F_h^{obs} - \langle F_h^{obs} \rangle)^2 \sum_h (F_h^{calc} - \langle F_h^{calc} \rangle)^2]^{1/2}$.

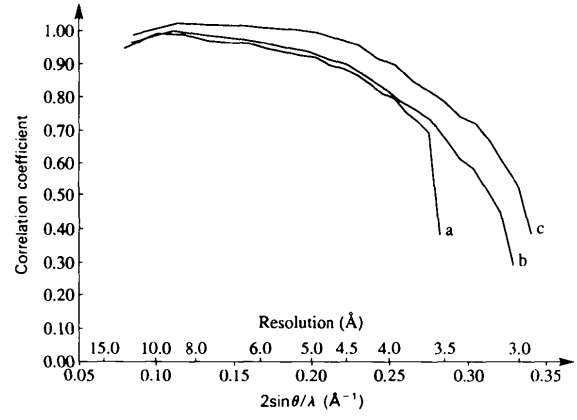

Fig. 6. Correlation coefficients (see Fig. 5 for definition) as a function of resolution. Curve *a*: Averaging with the 'old' data set between 15 and 3.5 Å resolution with a spherical mask. Curve *b*: Averaging with the 'old' data set between 15 and 3.0 Å resolution with a molecular mask. Curve *c*: Averaging with the 'new' data set between 15 and 2.9 Å resolution with the same molecular mask as used in *b*.

was calculated at 3.0 Å resolution. Its quality had improved, as compared to that at 3.5 Å resolution, in that more details such as clearer carbonyl bulges and flatter aromatic rings could be observed.

After the data had been reprocessed with the more accurate cell dimensions and scaled with resolution-dependent scale factors, a few more cycles of averaging were performed starting with the phases that had been obtained from the 'old' data set. The correlation coefficients improved quite dramatically for the entire resolution range (Fig. 6) with overall correlation coefficient of 0.89 (up from 0.83) and R factor of 16.1% (down from 21.2%). Both data quality and cell dimensions affect averaging. An inaccurate ratio between cell axes leads to averaging between incorrect grid points in the map. The map calculated with the new phase set showed better defined side chains. The phases were further extended to 2.9 Å resolution after incorporating calculated phases from the atomic model between 3.0 and 2.9 Å resolution. The difference between phases obtained from the 'old' and the 'new' data set increased steadily from a few degrees at low resolution to at least 30° at 3.5–3.0 Å resolution.

Refinement using the *PROLSQ* and *X-PLOR* programs

Virus orientation and position were first refined in a rigid-body least-squares procedure which minimizes the residual between observed structure-factor amplitudes and those calculated from the atomic model (Rossmann, 1976; Arnold & Rossmann, 1988). The orientation of the virus particle did not change significantly, but its position was moved by about 0.08 Å, from (0.5330, 0.5330, 0) to (0.5333, 0.5333, 0), which dropped the R factor by 0.2% for 1000 randomly selected reflections between 4.0 and 3.5 Å resolution. Later, this was repeated for the 'new' data set with the refined cell dimensions, but there was no further change in orientation or position of the particle.

Structure-factor phases were used in the structure refinement by *PROLSQ* to supplement the structure-factor amplitudes. Virus structure determination usually produces fairly accurate phases as a result of the phase restraints imposed by the high redundancy of non-crystallographic symmetry. The following residual

$$S_{\text{X-ray}} = \sum_h \{(1 - w_h)(|F_h^{\text{obs}}| - |F_h^{\text{calc}}|)^2 \quad \text{(scalar term)}$$

$$+ w_h[(A_h^{\text{obs}} - A_h^{\text{calc}})^2 + (B_h^{\text{obs}} - B_h^{\text{calc}})^2]\} \quad \text{(vector term)}$$

was minimized, where A_h and B_h are the real and imaginary components of the structure factor F_h, and w_h was the molecular replacement figure of merit mentioned above. Reflections with a good figure of merit contribute to the refinement primarily through the vector term, whereas reflections with a poor figure of merit contribute through the scalar term. This phase-refinement formulation was modified from the previous algorithm which contained vector terms alone (Arnold & Rossman, 1988). The current approach has the advantage that reflections with a low figure of merit (*e.g.* many high-resolution reflections) can still contribute to the refinement through scalar terms.

Non-crystallographically related viral subunits were constrained during refinement so that their positions and temperature factors obey icosahedral symmetry (Arnold & Rossmann, 1988; Krishnaswamy & Rossmann, 1990; Brünger, 1990). Contacts among icosahedrally related subunits were included in the non-bonded interactions, but the contacts between viral particles could not be restrained because constrained non-crystallographic symmetry refinement assumes the equivalence of all subunits.

Each round of *PROLSQ* refinement consisted of 20 to 30 cycles of conjugate-gradient least-squares minimization. Reflection subsets containing one-fourth of the total reflections, enough to over-determine the number of atomic parameters, were selected by taking every fourth reflection, in order to save computing time. This process is, therefore, reminiscent of Brünger's 'free' R factor (Brünger, 1992). To eliminate bias toward any one of the subsets, a new subset of reflections was used from cycle to cycle. Each subset gave very similar R factors (±0.2%). Partly refined atomic models were inspected after each round using *O* (Jones, Bergdoll & Kjeldgaard, 1990) and *FRODO* (Jones, 1978) with respect to the molecular-replacement averaged map.

The 'old' data set was used in the first three rounds of refinement (Table 3). Residues 156 to 161 were not included in the refinement because of their extremely weak electron density ($<0.5\sigma$). The electron density for the loop between residues 361 and 372 ($0.5\sigma <$ density $< 1\sigma$) showed possible alternative conformations. However, no attempt was made to build both conformations.

The slow-cooling molecular dynamics protocol in *X-PLOR* was used in the third round of refinement (Table 3). Because of the large amount of computation time that would be required, an initial temperature of 2000 K was used instead of the suggested 3000–4000 K (Brünger, Krukowski & Erickson, 1990) and the resolution of the reflection data was lowered to 3.2 Å. It has been shown that using 2000 K as an initial temperature might result in an R factor which is around 1% higher than that using 4000 K (Brünger *et al.*, 1990). All the reflections

Table 3. *Refinement results*

	Refinement with the 'old' data set				Refinement with the 'new' data set*		
	Round 0	Round 1	Round 2	Round 3	Round 4	Round 5	Round 6
Program	*PROLSQ*	*PROLSQ*	*PROLSQ*	*X-PLOR*	*PROLSQ*	*PROLSQ*	*PROLSQ*
Atomic model							
Residues	37–155 162–584	37–155 162–584	37–155 162 584	37–155 162–584	37–155 162–584	37–584	37–584 84 H_2O
Parameters refined		*x,y,z*	*x,y,z*	*x,y,z*	*x,y,z*	*x,y,z,B*	*x,y,z,B*
Data							
Resolution (Å)	5–3	5–3	5–3	6–3.2	6–3	6–3	6–3
F/σ cutoff	3.0	3.0	3.0	4.0	3.0	3.0	3.0
Reflections/cycle	45000	45000	45000	160000	45000	45000	45000
Bonding distance restraints							
1–2 distance (Å)†	0.025	0.020	0.018	0.030	0.021	0.022	0.020
1–3 distance (Å)‡	0.048	0.041	0.044	0.081	0.048	0.054	0.051
Bond angle (°)	1.8	1.7	2.4	4.8	2.4	2.9	2.9
1–4 distance (Å)§	0.080	0.063	0.050	0.111	0.062	0.077	0.072
Planarity restraint (Å^2)	0.023	0.018	0.013	0.040	0.012	0.007	0.007
Chiral volume restraint (Å^3)	0.149	0.163	0.192	0.161	0.086	0.061	0.062
Torsion-angle restraints							
Planar angle (°)	12.4	2.9	2.5	9.5	2.6	7.7	1.7
Staggered angle (°)	18.8	18.3	24.8	22.0	23.9	24.6	21.7
Orthogonal angle (°)	25.2	19.0	20.6	27.8	28.6	28.3	27.7
Non-bonded contact restraints							
1–4 contact (Å)	0.224	0.260	0.240	0.203	0.225	0.231	0.228
Contact other than 1–4 (Å)	0.397	0.342	0.328	0.207	0.270	0.230	0.265
Hydrogen-bonded contact (Å)	0.386	0.317	0.307	0.184	0.268	0.276	0.235
Thermal restraints							
1–2 main-chain motion (Å^2)						1.33	1.68
1–3 main-chain motion (Å^2)						2.10	2.85
1–2 side-chain motion (Å^2)						2.33	2.02
1–3 side-chain motion (Å^2)						3.46	3.39
R factors							
Scalar *R* (%)	34.8	28.5	28.3	25.4	24.1	22.0	21.1
Vector *R* (%)¶	55.7	39.8	39.3		36.0	31.5	29.7

* Including the more accurate cell dimensions.
† Bond length.
‡ Distance related to bond angles.
§ Distance related to fixed dihedral angles.
¶ $R_{\text{vector}} = \sum_h[(A_h^{\text{obs}} - A_h^{\text{calc}})^2 + (B_h^{\text{obs}} - B_h^{\text{calc}})^2]^{1/2}/\sum_h|F_h^{\text{obs}}|$.

above a 4σ cutoff between 6.0 and 3.2 Å resolution were used in the refinement. No phase restraint was applied. Two cycles of slow cooling, intervened by slight remodeling, gave a final *R* factor of 25.4% with an r.m.s. deviation in bond lengths of 0.030 Å and in bond angles of 4.8°.

The fourth to sixth rounds of refinement were performed with the 'new' data set and the refined cell dimensions (Table 3). Slight cell-dimension changes should be accommodated by slight distortion in geometry because the structure refinement was performed with respect to the atomic fractional coordinates. In the case of refinement with non-crystallographic symmetry constraints, however, the changes in cell dimension cannot be entirely accommodated because they affect the fractional positions of non-crystallographically related subunits.

In the fifth round of refinement, the restrained isotropic individual temperature factors were also refined in addition to the positional coordinates. Residues 156 to 161 were included and their temperature factors were initially assigned to 30 Å^2.

A difference Fourier map $(|F^{\text{obs}}| - |F^{\text{calc}}|)\exp(i\alpha_{\text{MR}})$ was calculated to locate solvent molecules, where F^{calc} were the structure factors of the atomic model after the fifth round of refinement and α_{MR} were the molecular replacement phases. Those peaks in the difference map that were higher than 3.5σ and had possible hydrogen-bonding neighbors were considered as putative water molecules. Water molecules that had moved during refinement to produce contacts of less than 2.5 Å with neighboring atoms or had temperature factors greater than 50 Å^2 were removed. A total of 84 water molecules were eventually kept in the final model. The final atomic model had an *R* factor of 21.1% with an r.m.s. deviation in bond lengths of 0.020 Å and in bond angles of 2.9°.

It has been shown (Rees & Lewis, 1983) that the phase error for reflections at 3 Å resolution should be less than about 30° in order to achieve a convergence for vector refinement that is faster than or comparable to scalar refinement. The average phase error was about 23° at the beginning of this refinement and 12° at the end of the refinement, as estimated from $(R_{\text{vector}} \times 180)/[\pi(2)^{1/2}]$ (Arnold & Rossmann, 1988). These errors are smaller than the allowable errors suggested above, although at the edge of the resolution, phase error can be considerably larger.

The structure

The overall structural features of CPV empty capsids are very similar to those described for CPV full

particles (Tsao *et al.*, 1991), except for the absence of electron densities of some ordered DNA and along the icosahedral fivefold axes. This result demonstrates that the quality of the data was sufficient for a successful structure determination in spite of the almost 800 Å long *c* axis.

Large conformational changes between the empty and full virus were observed at the interior surface of the virus around residues Cys490 to Cys494 where part of the ordered DNA binds in the CPV full-particle structure. The electron density along the fivefold axes in the monoclinic CPV full-particle structure had been interpreted to be the glycine-rich amino-terminal residues of the VP2 polypeptide (Tsao *et al.*, 1991). Although electron density along symmetry axes is prone to error, the existence of the density along the fivefold axes in CPV full particles was confirmed by a full minus empty particle difference map in the tetragonal space group using the 3.5 Å resolution phases of the CPV empty-particle structure. This confirms the observation using trypsin digestion that empty particles have the amino ends of VP2 inside, whereas in full particles some of the VP2 amino ends are externalized. A full analysis of the structure will be given elsewhere (Wu & Rossmann, 1993).*

We would like to thank Eli Lilly and Company for the use of their Cray 2S/2-128 computer and Jean-Pierre Wery for his help and assistance in running *X-PLOR* on Cray. We would also like to thank Mavis Agbandje, Jodi Bibler, Michael Chapman, Hok-Kin Choi, Vince Giranda, Vic Ilag, Kyung Kim, Prasanna Kolatkar, S. Krishnaswamy, Robert McKenna, Iwona Minor, M. R. N. Murthy, Marcos Oliveira, Andrew Prongay, Annette Shrive, Liang Tong, Jun Tsao, Peter Willingmann and Di Xia for help in the data collection of CPV at CHESS and also the CHESS staff for their assistance. We have appreciated useful discussions with Michael Chapman, Zhongguo Chen, John Johnson, Janet Smith and Liang Tong. We thank Helene Prongay and Sharon Wilder for help in the preparation of this manuscript. The work was supported by a Howard Hughes Medical Institute predoctoral fellowship to HW, an Austrian Science Foundation postdoctoral fellowship to WK, and grants from the National Science Foundation and the National Institutes of Health and a Lucille P. Markey Foundation Award to MGR.

* Atomic coordinates and structure factors have been deposited with the Protein Data Bank, Brookhaven National Laboratory (Reference: 1CAS, R1CASSF). Free copies may be obtained through The Technical Editor, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England (Supplementary Publication No. SUP 37092). A list of deposited data is given at the end of this issue.

References

ARGOS, P. & ROSSMANN, M. G. (1980). *Theory and Practice of Direct Methods in Crystallography*, edited by M. F. C. LADD & R. A. PALMER, pp. 361–417. New York: Plenum Press.

ARNOLD, E. & ROSSMANN, M. G. (1988). *Acta Cryst.* A**44**, 270–282.

BERNS, K. I. (1984). Editor. *The Parvoviruses*. New York: Plenum Press.

BRÜNGER, A. T. (1990). *X-PLOR Manual*. Yale Univ. Press, New Haven, CT, USA.

BRÜNGER, A. T. (1992). *Nature (London)*, **355**, 472–475.

BRÜNGER, A. T., KRUKOWSKI, A. & ERICKSON, J. W. (1990). *Acta Cryst.* A**46**, 585–593.

CLINTON, G. M. & HAYASHI, M. (1976). *Virology*, **74**, 57–63.

COTMORE, S. F. & TATTERSALL, P. (1987). *Adv. Virus Res.* **33**, 91–174.

HUBER, R. & KOPFMANN, G. (1969). *Acta Cryst.* A**25**, 143–152.

JOHNSON, J. E. (1978). *Acta Cryst.* B**34**, 576–577.

JONES, T. A. (1978). *J. Appl. Cryst.* **11**, 268–272.

JONES, T. A., BERGDOLL, M. & KJELDGAARD, M. (1990). *Crystallographic and Modeling Methods in Molecular Design*, edited by C. E. BUGG & S. E. EALICK, pp. 189–195. New York: Springer-Verlag.

KIM, S. (1989). *J. Appl. Cryst.* **22**, 53–60.

KRISHNASWAMY, S. & ROSSMANN, M. G. (1990). *J. Mol. Biol.* **211**, 803–844.

LUO, M., TSAO, J., ROSSMANN, M. G., BASAK, S. & COMPANS, R. W. (1988). *J. Mol. Biol.* **200**, 209–211.

MATTHEWS, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.

MATTHEWS, B. W. & CZERWINSKI, E. W. (1975). *Acta Cryst.* A**31**, 480–487.

PARADISO, P. R. (1981). *J. Virol.* **39**, 800–807.

PARADISO, P. R., RHODE, S. L. III & SINGER, I. I. (1982). *J. Gen. Virol.* **62**, 113–125.

PARRISH, C. R., AQUADRO, C. F., STRASSHEIM, M. L., EVERMANN, J. F., SGRO, J.-Y. & MOHAMMED, H. O. (1991). *J. Virol.* **65**, 6544–6552.

PARRISH, C. R., HAVE, P., FOREYT, W. J., EVERMANN, J. F., SENDA, M. & CARMICHAEL, L. E. (1988). *J. Gen. Virol.* **69**, 1111–1116.

RAYMENT, I. (1983). *Acta Cryst.* A**39**, 102–106.

REES, D. C. & LEWIS, M. (1983). *Acta Cryst.* A**39**, 94–97.

ROSSMANN, M. G. (1976). *Acta Cryst.* A**32**, 774–777.

ROSSMANN, M. G. (1979). *J. Appl. Cryst.* **12**, 225–238.

ROSSMANN, M. G. & BLOW, D. M. (1962). *Acta Cryst.* **15**, 24–31.

ROSSMANN, M. G., FORD, G. C., WATSON, H. C. & BANASZAK, L. J. (1972). *J. Mol. Biol.* **64**, 237–249.

ROSSMANN, M. G., LESLIE, A. G. W., ABDEL-MEGUID, S. S. & TSUKIHARA, T. (1979). *J. Appl. Cryst.* **12**, 570–581.

ROSSMANN, M. G., MCKENNA, R., TONG, L., XIA, D., DAI, J., WU, H., CHOI, H. K. & LYNCH, R. E. (1992). *J. Appl. Cryst.* **25**, 166–180.

SCHWAGER, P., BARTELS, K. & HUBER, R. (1973). *Acta Cryst.* A**29**, 291–295.

SIM, G. A. (1959). *Acta Cryst.* **12**, 813–815.

SIM, G. A. (1960). *Acta Cryst.* **13**, 511–512.

STUART, D. & WALKER, N. (1979). *Acta Cryst.* A**35**, 925–933.

TATTERSALL, P. & COTMORE, S. F. (1988). *Parvoviruses and Human Disease*, edited by J. R. PATTISON, pp. 5–41. Boca Raton: CRC Press.

TATTERSALL, P., SHATKIN, A. J. & WARD, D. C. (1977). *J. Mol. Biol.* **111**, 375–394.

TONG, L. & ROSSMANN, M. G. (1990). *Acta Cryst.* A**46**, 783–792.

TSAO, J., CHAPMAN, M. S., AGBANDJE, M., KELLER, W., SMITH, K., WU, H., LUO, M., SMITH, T. J., ROSSMANN, M. G., COMPANS, R. W. & PARRISH, C. R. (1991). *Science*, **251**, 1456–1464.

TSAO, J., CHAPMAN, M. S., WU, H., AGBANDJE, M., KELLER, W. & ROSSMANN, M. G. (1992). *Acta Cryst.* B**48**, 75–88.

WU, H. & ROSSMANN, M. G. (1993). *J. Mol. Biol.* In the press.