Acta Cryst. (1998). D**54**, 878–890

Structure Determination of the φX174 Closed Procapsid

TERJE DOKLAND,[a]† ROBERT MCKENNA,[a]‡ DEBRA M. SHERMAN,[b] BRIAN R. BOWMAN,[a]§ WILLIAM F. BEAN[a]¶ AND MICHAEL G. ROSSMANN[a]*

[a]*Department of Biological Sciences, Purdue University, West Lafayette, Indiana 47907, USA, and* [b]*Department of Botany and Plant Pathology, Purdue University, West Lafayette, Indiana 47907, USA.*
E-mail: mgr@indiana.bio.purdue.edu

(*Received 3 November 1997; accepted 9 February 1998*)

Abstract

The structure of a procapsid of the single-stranded DNA bacteriophage φX174 was determined to 3.5 Å resolution. The crystal space group was $I2_13$ with a unit-cell length of 774 Å. The unit cell contained 16 icosahedral virus particles, each situated on a crystallographic threefold axis. Thus, there are two independent one-thirds of a particle per asymmetric unit, and a total of 40-fold non-crystallographic redundancy. To aid in the interpretation of the packing arrangement, crystals were prepared for thin sectioning and analyzed by electron microscopy. Oscillation X-ray diffraction data was collected on image plates using synchrotron radiation and oscillation angles of either 0.25 or 0.30°. A low-resolution 6.5 Å data set collected from a single frozen crystal was particularly helpful in the structure determination, because of its completeness and internal consistency. The initial particle orientations were determined using self-rotation functions, while the initial position of one particle was determined from a Patterson map. The structure was solved by molecular replacement real-space averaging using a model based on a cryo-electron microscopy reconstruction as a starting point for the phase determination. The initial structure determination used the data between 20 and 13 Å resolution, which was then extended one reciprocal lattice point at a time to 6.5 Å resolution. At this point, a 3.5 Å resolution data set compiled from a number of crystals collected at 277 K was introduced. Phase extension and averaging continued to 3.5 Å resolution after re-determining the particle positions and orientations. The amino-acid sequences of most of the D, F and G proteins and part of the B protein could be unambiguously built into the 3.5 Å electron-density map. Partial crystallographic refinement yielded an *R* factor of 31.6%, consistent with the relatively low resolution and lack of completeness of the data.

1. Introduction

The bacteriophage φX174 is a single-stranded DNA virus of *Escherichia coli* (Hayashi *et al.* 1988). It has an icosahedral capsid of characteristic shape with protruding 'spikes' on the fivefold axes and a diameter of approximately 260 Å between twofold vertices or 330 Å between fivefold vertices (Hall *et al.*, 1959; Sinsheimer, 1959; McKenna, Xia, Willingmann, Ilag, Krishnaswamy *et al.*, 1992). The mature capsid contains 60 copies each of the capsid protein F (48.4 kDa), the spike protein G (16.9 kDa) and the DNA-associated protein J (4.2 kDa), as well as 12 copies of the 'pilot' protein H (34.4 kDa) (Burgess, 1969).

During assembly, pentamers of the F and G proteins (9S and 6S particles, respectively) form spontaneously inside the infected cells (Tonegawa & Hayashi, 1970). The two scaffolding proteins D and B are required for the assembly to proceed further (Siden & Hayashi, 1974; Fujisawa & Hayashi, 1977*b*). In the presence of scaffolding proteins, a precursor capsid, called the procapsid or 108S particle, is assembled (Fujisawa & Hayashi, 1977*a*; Mukai *et al.*, 1979). The procapsid contains 240 copies of the D protein forming an external scaffold, and 60 copies of the internally located B protein, in addition to the F, G and H proteins. The genome is packaged into the procapsid together with the DNA-associated protein J, at which time the B protein exits. The resulting 132S intermediate retains the outer scaffold. Removal of the scaffold, which may be associated with increased Ca^{2+} concentration encountered upon host cell lysis, yields the mature infectious 114S virion (Weisbeek & Sinsheimer, 1974; Fujisawa & Hayashi, 1977*c*).

The structure of the virion at 3.0 Å resolution was reported previously (McKenna, Xia, Willingmann, Ilag, Krishnaswamy *et al.*, 1992; McKenna *et al.*, 1994). Recently, we reported the structure of a scaffold-containing procapsid particle (Dokland *et al.*, 1997) which for the first time shows a scaffolding protein structure at high resolution. These proteins are of great

† Present address: Institute of Molecular Agrobiology, The National University of Singapore, Singapore 117604.
‡ Present address: Department of Biological Sciences, University of Warwick, Coventry CV4 7AL, UK.
§ Present address: Department of Biochemistry, Baylor College of Medicine, Houston, Texas 77030, USA.
¶ Present address: Indiana University School of Medicine, Indianapolis, Indiana 46202, USA.

importance in viral assembly and may be essential components of many other macromolecular assembly systems. The function of these proteins is to catalyze the assembly of a macromolecular complex, not unlike chaperones, which direct the folding of proteins. Based on comparisons with the previously determined 25 Å resolution electron microscopy (EM) structure (Ilag *et al.*, 1995), we concluded that, during crystallization, this procapsid had gone through part of the maturation process normally associated with DNA packaging. The X-ray structure was termed the 'closed procapsid', because openings at the threefold axes seen in the EM reconstruction (the 'open procapsid') were absent. In this paper, we report the crystallization and structure determination of the closed procapsid. The large size of the cubic unit cell (a = 774 Å) and the presence of two independent particles posed problems in the structure determination. The analysis used EM images of thin-sectioned crystals to aid in the interpretation of the crystal lattice. The X-ray structure of the closed procapsid was then solved using molecular replacement with the EM reconstruction of the open procapsid as a starting model.

2. Procapsid purification and crystallization

Originally, the purification protocol outlined in Ilag *et al.* (1995) was followed. In this protocol, lysis-defective φX174 *am(E)W4* was used to infect non-suppressing *E. coli* C using a standard protocol (Incardona *et al.*, 1985). Both virions and procapsids are produced in such infections. The cells were lysed by treatment with lysozyme in 0.1 *M* borate buffer, pH 9.0, containing 30 m*M* EDTA, and three cycles of freezing/thawing. After removal of debris by centrifugation at 25 000*g* for 15 min, PEG 8000 was added to a final concentration of 5%, and the solution stirred overnight at 277 K and then centrifuged at 25 000*g* for 15 min. Under these conditions, virions collected in the pellet, while the supernatant contained the procapsids. The procapsids were then collected by centrifugation at 270 000*g* for 60 min and purified either on a 10–40% potassium tartrate gradient in 0.1 *M* borate or on a 5–25% sucrose gradient in 0.1 *M* borate containing 5% potassium tartrate. The tartrate had previously been shown to enhance the stability of the procapsid for prolonged periods of time (Ilag *et al.*, 1995).

More recently, a different protocol was developed (Dokland *et al.*, 1997), which increased the yield and improved the consistency of the purification. In this method, φX174 *am(E)W4* was used to infect *E. coli* gro89 cells which are deficient in viral DNA synthesis, resulting in the arrest of viral assembly prior to DNA packaging (Ekechukwu *et al.*, 1995) and hence the accumulation of procapsids in the host cells. The purification procedure essentially followed that of Mukai *et al.* (1979). The cells were lysed by lysozyme treatment in buffer *A* [0.1 *M* tris-(hydroxymethyl) amino methane–HCl (Tris–HCl), pH 7.5, 0.1 *M* NaCl and 25 m*M* EDTA], and three cycles of freezing/thawing. After removal of debris by centrifugation at 13 000*g* for 20 min, the supernatant was precipitated with 40%(*v*/*v*) (final concentration) saturated ammonium sulfate in buffer *A* on ice for 1 h. After centrifugation at 18 000*g* for 20 min, the pellets were resuspended in 0.02 *M* Tris–HCl, pH 7.5, containing 1 m*M* $MgCl_2$ and 0.05 *M* NaCl, dialyzed overnight against buffer *A* and then purified twice on 10–40% potassium tartrate gradients in 0.02 *M* Tris–HCl, pH 7.5, containing 5 m*M* EDTA, 1 m*M* $MgCl_2$ and 0.1 *M* NaCl. The purified material produced by either method was finally concentrated by centrifugation at 23 000*g* for 90 min and resuspended in a sample buffer containing 10 m*M* Tris–HCl, pH 7.5, 1 m*M* EDTA, 1 m*M* $MgCl_2$ and 0.25 *M* NaCl, to a final procapsid concentration of 15–20 mg ml^{-1}.

The φX174 procapsids were crystallized using hanging-drop vapour diffusion (Weber, 1997). The reservoir contained 200 µl of 0.5 *M* 2-(*N*-morpholino) ethane sulfonic acid (MES), pH 6.0, and 430–470 µl saturated ammonium sulfate to a final volume of 1 ml [43–47%(*v*/*v*) ammonium sulfate]. Usually, 5 µl of procapsid solution in sample buffer was mixed with 5 µl of the reservoir solution. The crystallization was carried out at 277 K. Crystals typically appeared after 2–3 weeks, but sometimes only after several months. The crystals were extremely sensitive to temperature variations and had to be kept at 277 K during storage, transportation and data collection. The crystals were commonly 0.1–0.2 mm in size, but sometimes grew larger than 0.5 mm. The crystals produced from procapsids purified by the two methods were isomorphous, and X-ray data was collected at 277 K from crystals produced by both methods.

Various cryo-protectant solutions were tried for freezing the crystals (Rodgers, 1997). The most success was achieved using 20–25% glycerol. Crystals could be grown with 20% glycerol present in the reservoir solution, or were soaked for 5–60 min in glycerol-containing stabilizing solution [20% glycerol, 10 m*M* MES, pH 6.0, 47–50%(*v*/*v*) saturated ammonium sulfate] prior to freezing. However, even in this solution the crystals were only partly stable, and significant disorder was observed. For freezing, the crystals were picked out of the cryo-protectant solution with a nylon loop, plunged into liquid ethane slush at about 123 K and stored under liquid nitrogen until data collection. Only crystals produced using the second method of purification were used to collect low-resolution data at liquid-nitrogen temperatures.

3. X-ray diffraction data collection and processing

The diffraction data was collected using oscillation photography, with oscillation angles of either 0.25 or

Table 1. *Data-collection statistics*

Frozen data set (1 crystal)

Resolution (Å)	R_{merge}† ($R_{merge,2\sigma}$)	$\langle I/\sigma(I)\rangle$	Completeness‡ (Completeness,2σ)
∞–12.97	7.5 (5.4)	20.8	95.2 (77.9)
12.97–10.31	7.3 (5.3)	21.5	96.4 (80.7)
10.31–9.01	8.9 (6.0)	18.0	96.9 (77.1)
9.01–8.19	12.8 (7.3)	12.9	97.1 (70.1)
8.19–7.60	22.2 (9.7)	7.4	97.4 (57.7)
7.60–7.15	33.6 (12.2)	5.0	97.6 (49.0)
7.15–6.80	44.6 (13.6)	3.6	97.7 (41.4)
6.80–6.50	54.1 (15.9)	3.0	97.9 (38.3)
All	15.0 (6.7)	10.8	97.0 (61.6)

Warm data set (22 crystals)

Resolution (Å)	R_{merge}†	$\langle I/\sigma(I)\rangle$	Completeness‡
∞–10.78	9.8	5.0	83.1
10.78–7.73	12.1	4.8	84.3
7.73–6.35	21.8	3.0	85.4
6.35–5.51	31.8	2.0	84.1
5.51–4.93	33.7	1.9	81.5
4.93–4.51	38.3	1.6	73.4
4.51–4.18	41.0	1.4	54.5
4.18–3.91	46.5	1.3	41.5
3.91–3.69	37.8	1.6	25.7
3.69–3.50	40.4	1.5	13.5
All	24.7		55.3

† $R_{merge} = 100 \times \sum_h \sum_j |I_{hj} - \langle I_h\rangle| / \sum \sum I_{hj}$; $R_{merge,2\sigma}$ is calculated using only data for which $I > 2\sigma(I)$. ‡ Completeness = 100 × (No. observed unique reflections)/(No. possible unique reflections); for the frozen data, figures in parentheses are calculated using only data for which $I > 2\sigma(I)$.

0.30° and crystal-to-detector distances varying from 350 to 600 mm. Most of the images were recorded on FUJI image plates at the Cornell High Energy Synchrotron Source (CHESS) beamline F-1, using 0.91 Å wavelength radiation. A few images were recorded with a MAR 30 cm image-plate scanner at the Brookhaven National Synchrotron Light Source (NSLS) beamline X12c, using 1.10 Å radiation.

A 97% complete data set to 6.5 Å resolution (Table 1) was collected at about 103 K from a single frozen crystal (the 'frozen data'). Although the frozen crystals never diffracted to beyond about 6.0 Å resolution due to disorder introduced by the glycerol soaking, the frozen data set did offer a good starting point for the structure determination due to its high completeness and redundancy. Furthermore, as the frozen data set came from a single crystal, we were assured of internal consistency and no problems associated with the combination of data from non-isomorphous crystals (see below). To obtain higher resolution data, however, it was necessary to combine data from a number of crystals collected at 277 K (the 'warm data'). Due to the high radiation sensitivity of the warm crystals, typically only 2–4 images could be recorded from each crystal. In addition, only some crystals gave good diffraction beyond 4.5 Å resolution. Thus, the completeness of this data dropped dramatically at higher resolution (Table 1). Furthermore, the warm data is extremely weak, with an average $I/\sigma(I)$ of less than 2.0 beyond 5.5 Å.

The diffraction images were indexed and processed using the program *Denzo* (Otwinowski & Minor, 1997). The crystals belonged to a cubic body-centred space group with Laue group $m3$. The average cell dimension for the warm crystals was a = 774 Å, while the frozen crystal had a unit-cell length of a = 769 Å, consistent with a shrinkage of the cell due to the freezing. Out of the two possible space groups $I23$ or $I2_13$ later analysis (see below) showed the correct space group to be $I2_13$. These space groups allow two alternative axial settings related by a 90° rotation to each other. R_{merge} was used to differentiate the two settings relative to an already consistent reference data set. Because of the poor quality of the data, it was often not possible to distinguish between the two settings. This ambiguity existed for more than half of all the crystals, and this data was therefore rejected and not included in the final warm data set.

The frozen data, which consisted of a continuous series of 125 oscillation images, could be easily scaled and merged by adding together partial reflections from abutting images, using the program *SCALEPACK* (Otwinowski & Minor, 1997). The warm data, on the other hand, consisted mostly of sets of 1–4 images located at random in reciprocal space. As the mosaicity was often relatively high compared with the oscillation range, most of the warm data consisted of unpaired partial reflections. Since *SCALEPACK* would normally reject these reflections, an alternative procedure was used in order to avoid losing too much of the data. In this scheme, *SCALEPACK* was used initially for the scaling and formatting of the data, but not for merging. The reflections were then re-scaled using the program *SCALA* (Evans, 1993) of the *CCP*4 suite (Collaborative Computational Project, Number 4, 1994). Reflections that could not be added to create a full reflection were rejected if the calculated partiality was less than 0.5 or otherwise corrected by the calculated partiality. The data was then merged using *AGROVATA* in the *CCP*4 suite. No cutoff of weak reflections was used during this procedure, but the sigmas of all the reflections in each intensity range were corrected to reflect the scatter of observations within that intensity range (Rossmann *et al.*, 1979; Evans, 1993). Finally, the reflection intensities of both the frozen and warm data were converted to amplitudes using the program *TRUNCATE* in the *CCP*4 suite, which utilizes maximum-likelihood methods to deal with negative and very weak diffraction intensities (French & Wilson, 1978).

4. Packing considerations

Using a molecular weight of 9.3×10^6 Da for the procapsid particle, the Matthews coefficient (V_M) is 6.2 Å^3 Da^{-1} when there are eight particles in the unit

cell, 4.1 Å^3 Da^{-1} for 12 particles and 3.1 Å^3 Da^{-1} for 16 particles per unit cell. These are the only possible numbers of particles consistent with the space-group symmetry (either $I23$ or $I2_13$) which give reasonable V_M values. With eight particles per unit cell, each particle would have to occupy a special position on the body diagonal threefold axis in either space group. The closest packing arrangement would then leave the particles separated by 387 Å, larger than the anticipated diameter of the particle based on EM studies (Ilag *et al.*, 1995). If there were 12 particles per unit cell, they would have to be situated on crystallographic twofold axes, which would not be possible based on packing considerations. Placing 16 particles in the unit cell in space group $I2_13$ would require two independent particles per unit cell, each sitting on a crystallographic threefold body diagonal. With the independent particles at approximately (0,0,0) and $(\frac{1}{4},\frac{1}{4},\frac{1}{4})$ positions, a close-packed arrangement would be achieved with a closest approach of 335 Å, somewhat less than the mean diameter of 360 Å determined by EM (Ilag *et al.*, 1995). The particles could not be situated exactly at (0,0,0) and $(\frac{1}{4},\frac{1}{4},\frac{1}{4})$, however, since the unit cell would then break down to a smaller cell with half the unit-cell length. A closely similar arrangement would also be possible in space group $I23$ with eight symmetry-related particles near $(\frac{1}{4},\frac{1}{4},\frac{1}{4})$ on threefold axes, six particles at special 222 positions and two particles at special 23 positions.

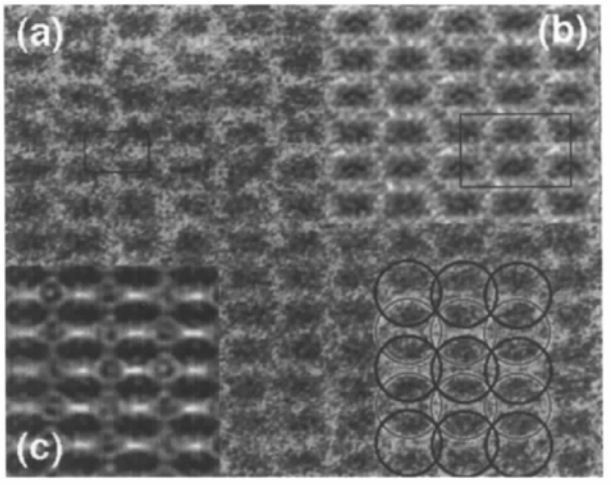


Fig. 1. (*a*) Electron micrograph of a thin section through a φX174 procapsid crystal. The apparent unit cell of 260 × 250 Å is indicated. (*b*) The same thin section electron micrograph Fourier filtered to 55 Å resolution. The circles represent the arrangement of procapsid shells in the lattice. Overlap between the empty procapsid shells gives rise to the pattern of dark bands in the micrograph. The rectangular outline which is indicated has dimensions 720 × 500 Å, corresponding to the crystallographic unit cell projected at a 45° angle (see Fig. 2). (*c*) Projection of a model of a unit cell containing 16 procapsid particles. The unit cell was projected down the [01−1] axis, which is at a 45° angle to one of the principal axes of the cell. Note the similarity of the resulting overlap pattern with that observed in the sectioned crystal.

5. Thin sectioning and electron microscopy

In the previous analysis of the G4 'degraded procapsid', the procapsid had lost its outer scaffold during crystallization, leaving a structure essentially identical to the virion (McKenna *et al.*, 1996). To ascertain that this had

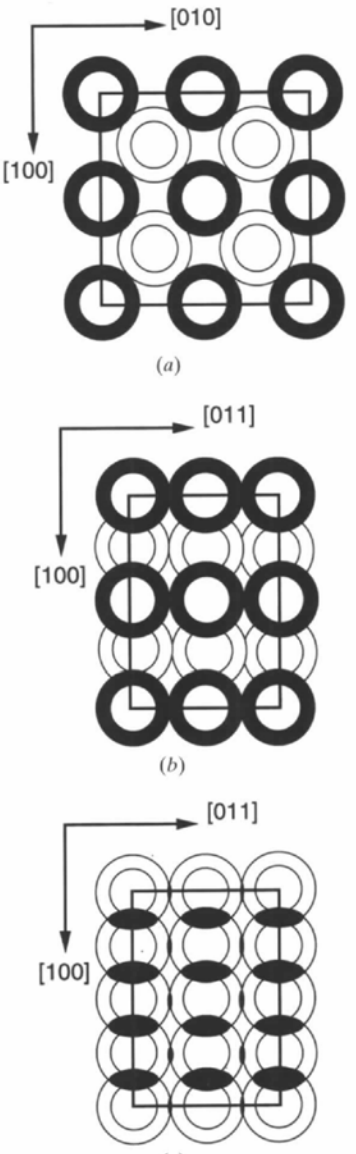


Fig. 2. Interpretation of the crystal thin sections. Each procapsid particle is shown as a ring, representing the electron-dense shell of the capsid with its empty interior. (*a*) A view down one of the principal axes [001] of the crystal unit cell shows the closely packed arrangement of interspersed layers of particles that would ensue with 16 procapsid particles in the unit cell. (*b*) Viewing the unit cell at an angle of 45° (down the [01−1] axis) brings some of the particles into register. Projection of the particles down this axis results in an appearance like that in (*c*), due to the superposition of the dense shells. This pattern corresponds to the observed projection in the micrographs (Fig. 1).

not happened in the φX174 procapsid crystals, some crystals were prepared for thin sectioning and observation by EM (Johnson *et al.*, 1974). This analysis also gave an indication of the packing arrangement in the crystals.

The crystals were washed in stabilizing solution [47%(*v*/*v*) saturated ammonium sulfate, 0.1 *M* MES, pH 6.0] and fixed in 0.1% glutaraldehyde in stabilizing solution at 277 K for 72 h. After washing briefly with 0.1% glutaraldehyde in 0.067 *M* phosphate buffer, they were post-fixed in 1.0% OsO_4 in 0.067 *M* phosphate for 1 h. After a brief rinse in phosphate buffer, the crystals were dehydrated in a graded ethanol series (70–100%), followed by 100% propylene oxide. After a stepwise infiltration in propylene oxide/LX-112 resin (LADD Research Industries, Burlington, VT) for a total of 48 h, the crystals were embedded in LX-112 resin in Beem capsules and polymerized at 333 K for 48 h. Silver sections were cut, placed on 300 mesh grids and post-stained with 2.0% uranyl acetate followed by Reynolds lead citrate (Reynolds, 1963). The samples were examined at a magnification of 46 000× or 60 000× with a Philips 400 transmission electron microscope. The negatives were digitized on an Optronics rotating-drum scanner and analyzed with image-analysis software provided by Tim Baker (Purdue).

The micrographs showed a pattern of dark rectangles with a spacing of about 360 Å in the *x* direction and 250 Å in y (Fig. 1*a*). Fourier analysis and filtration of the repeat pattern revealed a unit cell of 720 × 500 Å with a pattern of alternating inverted blocks of density (Fig. 1*b*). These dimensions correspond approximately to the unit-cell length and one-half of a face diagonal length (770 and 535 Å, respectively), indicating that the crystal had been sectioned perpendicular to the [0 1 1] axis, which is 45° to one of the principal axes in the crystal (Fig. 2). The reason for this is that the crystal has its largest face parallel to a face diagonal. Apparently, on preparation for thin sectioning, this face had oriented itself perpendicular to the sectioning axis (0 1 −1).

Since the EM section is about 800 Å thick, the micrograph represents a superposition of several layers of particles. It was, therefore, difficult to interpret the observed pattern directly in terms of spherical virus particles. However, packing considerations, as well as analysis of the diffraction data (see below), limited the number of possible particle arrangements to the two

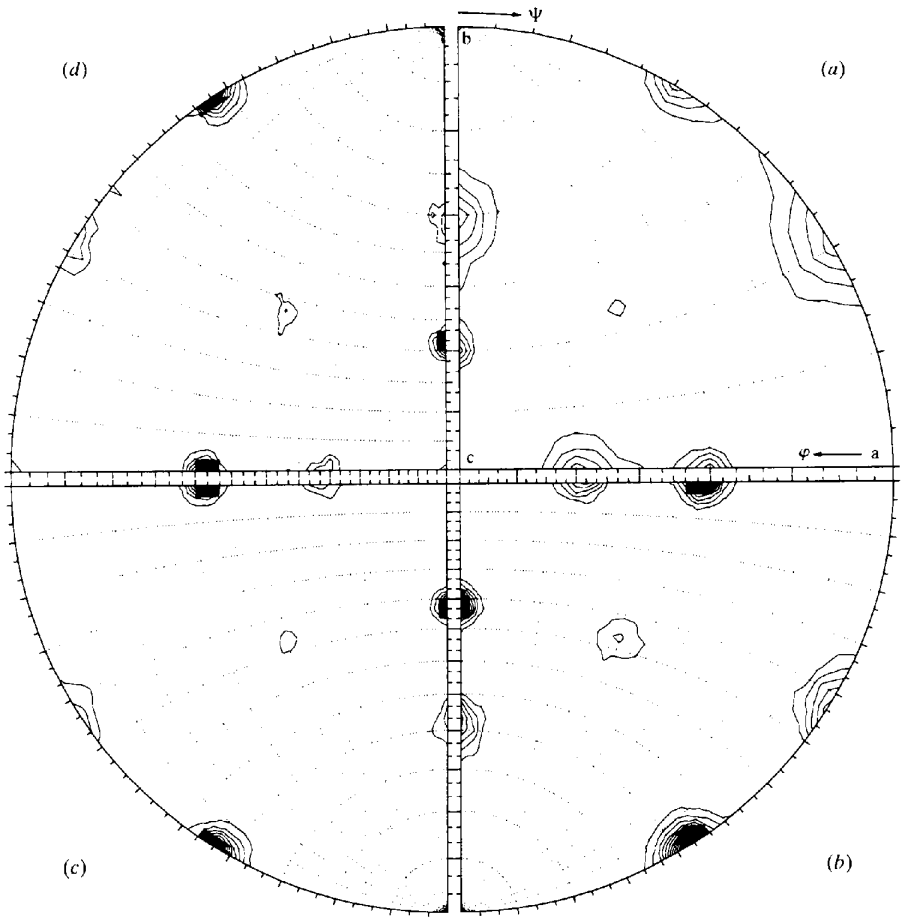


Fig. 3. Rotation-function search for fivefold NCS axes (κ = 72°) of the warm data set in different resolution ranges. Clockwise, from (*a*) between 45 and 15 Å resolution; (*b*) 15–8 Å resolution; (*c*) 8–6 Å resolution; (*d*) 6–5.5 Å resolution. Note how one set of peaks gets weaker at higher resolution.

possibilities of either eight or 16 particles per unit cell. Consequently, models were made by placing either eight or 16 copies of the EM reconstruction of either the procapsid or the virion into the unit cell, and projecting this model down the [0 1 −1] axis. This showed that only a set of 16 procapsid-shaped particles was consistent with the observed pattern in the micrographs (Fig. 1*c*). The observed pattern of dark rectangles results from the superposition of the edges of the closely packed empty procapsid shells, which, when the unit cell is projected down [0 1 −1], are brought into register (Fig. 2). This analysis thus showed that, indeed, the crystals contained procapsids, and not empty virion-like particles. Furthermore, it strongly suggested that there were 16 particles in the unit cell in a cubic close-packed arrangement.

6. Determination of particle positions and orientations

Self-rotation functions were calculated in various resolution ranges using the program *GLRF* (Tong & Rossmann, 1990) with a radius of integration of 180 Å. Searches for peaks were made on the κ = 72, 120 and 180° planes, which locate local five-, three- and twofold axes in the unit cell, respectively. Using data to about 25 Å resolution, two sets of icosahedrally distributed peaks were found, rotated relative to each other by 90° (Fig. 3), consistent with the presence of 16 particles in the unit cell. Both sets had their icosahedral twofold axes nearly parallel to the crystallographic twofolds. At higher resolution, however, one set of icosahedral peaks became progressively weaker relative to the other set (Fig. 3). This phenomenon suggested that perhaps one set of peaks was an artefact. However, this was not a very satisfactory explanation, because of the poor packing arrangement that would ensue with only eight particles in the unit cell. Furthermore, rotation functions based on models with eight particles per unit cell did not generate any such artefactual peaks. In addition, the crystal thin-sectioning results had strongly suggested that there were 16 particles in the unit cell.

At this stage, we were working with the older warm data set, which was generated from a number of crystals, as we had not yet successfully frozen crystals. We were, therefore, considering possibilities such as that the many crystals comprising the warm data were somehow different to each other, that the individual crystal data had been incorrectly merged, or that the crystals had some kind of internal 'twinning', each crystal consisting of a mosaic of lattices in different orientations. (The latter might still be the case for those crystals that could not be merged to the rest of the data set.) When the first frozen data set, collected from a single crystal, behaved in the same way as the warm data, however, it was clear that the disappearance of one set of rotation-function peaks at higher resolution was not due to differences between crystals and had to be a phenomenon internal to the individual crystal. It turned out later that the disappearance of the peaks was due to one of the particles being rotated by about 2° away from the special setting for which the orthogonal icosahedral twofold axes are aligned with the crystallographic twofold axes. This leads to a fourfold splitting of the peaks at higher resolution, and hence a reduction in peak intensity.

The presence of non-crystallographic twofold axes parallel to within a few degrees of the crystallographic twofold axes should give rise to very large peaks in the Patterson function (Tsao, Chapman, Wu *et al.*, 1992). These peaks can identify the position of the non-crystallographic twofold axes, corresponding to the centres of the particles, in the unit cell. In space group $I23$ such peaks would be expected to be at $(2x,2x,0)$, $(2x,0,2x)$ and $(0,2x,2x)$, while in $I2_13$ they would appear at $(0,2x,\frac{1}{2}-2x)$, $(\frac{1}{2}-2x,0,2x)$ and $(2x,\frac{1}{2}-2x,0)$. A Patterson function calculated using the data between 15 and 8 Å resolution showed strong peaks at u,v,w = (0.49,0,0.01) and all its symmetry-related equivalent positions (Fig. 4). This therefore established that the space group was $I2_13$ and that one icosahedral particle has its centre at x,y,z = (0.245,0.245,0.245), *i.e.* close to the $(\frac{1}{4}, \frac{1}{4}, \frac{1}{4})$ position. There was no sign of the second independent particle, for which the rotation-function peaks started to fade at higher resolution than about 15 Å. The reason that the second particle was not detected in the Patterson map was that the 2° rotation that had caused the

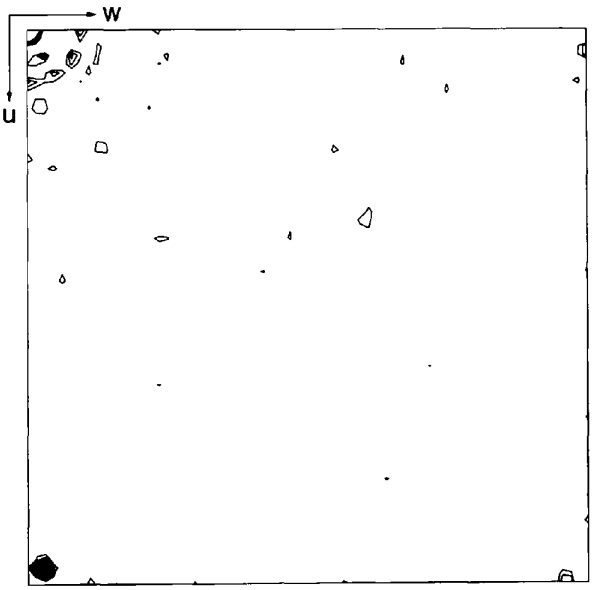


Fig. 4. Harker section (v = 0) of the Patterson map calculated between 45 and 15 Å resolution. One very strong peak is apparent at u,v,w = (0.49,0,0.01), consistent with a particle with local twofold axes aligned with the crystallographic twofold axes, located at x,y,z = (0.245,0.245,0.245) in space group $I2_13$. Contours start at 2σ and go in steps of 2σ.

Table 2. *Molecular-replacement averaging statistics*

Data set	Averaging cycle No.	Particle positions†				Resolution (Å)	CC‡	*R*§
		1		2				
		κ	*x*	κ	*x*			
Frozen data, one particle¶								
	A0	0.00	0.2450	–	–	13.0	0.147	65.7
	A1	0.00	0.2450	–	–	13.0	0.392	53.1
	A31	0.00	0.2448	–	–	13.0	0.597	42.6
Frozen data, two particles								
	B0	0.00	0.2448	−43.40	−0.0050	13.0	0.484	48.6
	B1	0.00	0.2448	−43.40	−0.0050	13.0	0.582	43.8
	B20	0.00	0.2448	−43.40	−0.0050	13.0	0.689	36.6
	B48	−0.10	0.2448	−43.36	−0.0047	13.0	0.701	35.4
	B260	−0.62	0.2446	−42.69	−0.0046	9.4	0.942	14.0
	B447	−0.63	0.2446	−42.68	−0.0046	6.5	0.956	13.2
Warm data, two particles								
	C1	−0.63	0.2446	−42.68	−0.0046	9.0	0.677	35.1
	C305	−0.04	0.2432	−41.31	−0.0052	4.0	0.927	16.4
	C450	−0.03	0.2432	−41.29	−0.0052	3.5	0.868	23.3

† κ is the rotation around the body diagonal threefold axis of the cubic unit cell which relates an icosahedron in the standard orientation with orthogonal twofold axes on the *x*, *y* and *z* axes (Tsao, Chapman, Wu *et al.*, 1992) to the orientation in the crystal. The orientation of the body diagonal is given by polar angles $\psi = 54.7$ and $\varphi = -45.0$ (Rossmann & Blow, 1962). The polar angles give rise to a [**P**] matrix that relates the coordinates (**X**′) of the standard icosahedron to the coordinates **X** in the unit cell such that $\mathbf{X} = [\mathbf{P}]\mathbf{X}' + \mathbf{t}$ where **t** is the translation vector $\mathbf{t} = (t_1, t_2, t_3)$.

$$\text{For particle 1, } [\mathbf{P}] = \begin{bmatrix} 1.00000 & 0.00030 & -0.00030 \\ -0.00030 & 1.00000 & 0.00000 \\ 0.00030 & -0.00030 & 1.00000 \end{bmatrix}, \quad \text{for particle 2, } [\mathbf{P}] = \begin{bmatrix} 0.83425 & 0.46385 & -0.29810 \\ -0.29810 & 0.83425 & 0.46385 \\ 0.46385 & -0.29810 & 0.83425 \end{bmatrix}.$$

For particle 1, **t** = (188.082, 188.082, 188.082) and for particle 2, **t** = (−4.0248, −4.0248, −4.0248). The position of the particle in the unit cell is given by the fractional coordinates $x = t_1/a (x = y = z)$. ‡ $CC = [\sum_h (\langle F_{obs}\rangle - F_{obs,h})(\langle F_{calc}\rangle - F_{calc,h})]/[\sum_h (\langle F_{obs}\rangle - F_{obs,h})^2 \times \sum_h (\langle F_{calc}\rangle - F_{calc,h})^2]^{1/2}$ § $R = 100 \times \sum_h |F_{calc,h} - kF_{obs,h}| / \sum_h |F_{calc,h}|$, where $k = \langle F_{calc}\rangle / \langle F_{obs}\rangle$ for all reflections within each resolution bin. ¶ Initially, only the particle at $(\frac{1}{4},\frac{1}{4},\frac{1}{4})$ was used for phasing. The particle at (0,0,0) was introduced after cycle A31.

disappearance of the rotation-function peaks at higher resolution also led to a failure to produce detectable peaks.

7. Molecular replacement

Molecular replacement followed the procedure described by Rossmann (1990) and Arnold *et al.* (1987). There are two independent thirds of a particle in the crystallographic asymmetric unit, or two independent sets of eight particles in the unit cell. As the particles are sitting on crystallographic threefold axes, each of the two independent particles has 20-fold non-crystallographic symmetry (NCS). The program *ENVELOPE* (Rossmann *et al.*, 1992) was used for the real-space electron-density averaging. A simple, spherical shell with an outer radius of 180 Å and an inner radius of 75 Å was used as a mask, outside of which the density was flattened. There is one orientational (rotation about the threefold axis) and one positional (translation along the threefold axis) parameter that have to be determined for each independent particle. The rotation function and the Patterson had showed that only one of the two particles was closely aligned with the crystallographic twofold axes; therefore, the position of only one of the particles was well defined at the beginning of molecular replacement (Table 2).

The previously determined EM reconstruction (Ilag *et al.*, 1995) was used as a starting model for the phase determination. The reconstruction, however, had a resolution of only about 25 Å, while X-ray data was missing inside about 45 Å resolution due to the presence of the beam stop. To bridge the resolution gap between the EM and X-ray data, a model was generated in which the known X-ray structure of the virion (McKenna, Xia, Willingmann, Ilag, Krishnaswamy *et al.*, 1992) was fitted into the EM density. This required extensive modification of the virion structure (Ilag *et al.*, 1995). For those parts of the structure corresponding to the external and internal scaffolding proteins, for which no X-ray structure was available, the G spike protein was used as a model, as it had approximately the right molecular weight and a compact, rounded shape. Structure factors were then calculated to 13 Å resolution. These were used to compute an electron-density map consisting of a single particle in the standard orientation. This density was then placed into the crystallographic unit cell using the previously determined positions and orientations (Fig. 5).

Initially, only the first particle was used for the phase determination, in conjunction with the frozen data set, taking the data between 20 and 13 Å resolution. This particle was placed at $x = y = z = 0.245$ with an orientation $\kappa = 0.0$. After 31 cycles of electron-density averaging, the correlation increased from 0.15 to 0.60 (Table 2). At this point, the second particle was introduced, using the currently averaged particle as a new model. The position of the second particle was inferred from packing considerations to be at $x = y = z = -0.005$, or translated by exactly $\frac{1}{4}$ unit-cell length away from the first particle. The orientation was taken to be 90° rotated from the first particle as a first approximation. A 90° rotation around the local twofold axis corresponds to a rotation of $\kappa = -44.48°$ around the threefold axis (Table 2). This starting model was optimized by a 'coarse grid' search procedure in which the orientation and position of the second particle was varied within the unit cell that already contained the first particle. For each change in one of these parameters, the correlation was calculated between the observed structure factors and those derived from this map. The initial correlation after this search procedure was 0.48 (Table 2). After 20 cycles of averaging, the correlation had increased to 0.69. Subsequently, the resolution was extended one reciprocal lattice interval at a time to 9 Å resolution. After 260 cycles of averaging, we noticed a drop in the correlation between F_{calc} and F_{obs} at around 15 Å resolution. It was assumed that the phases were inverted ($\alpha' = \alpha + \pi$) between 20 and 15 Å resolution compared with the higher resolution data (see below). Consequently, the resolution range was narrowed to 14.5–9 Å, and phase extension proceeded one reciprocal lattice interval at a time in both directions until all the data between 23 and 6.5 Å had been included. Every ten cycles or so during the phase extension the positions and orientations of the two particles were independently refined using the 'climb' procedure (Muckelbauer *et al.*, 1995). In this procedure, the r.m.s. deviation between NCS-related points is minimized while moving the particle positions and orientations in the unit cell. After each round of climbing, another ten cycles of averaging were performed. Typically, the orientation of each particle changed by less than 0.05° per climb cycle, but would continue changing after each round of ten averaging cycles, until both particles had moved by about 0.7° and 0.5 Å. Once the orientations and positions had stabilized at 9 Å resolution, they did not move any further during the subsequent phase extension. The final correlation for the frozen data at 6.5 Å resolution was 0.96 (Table 2).

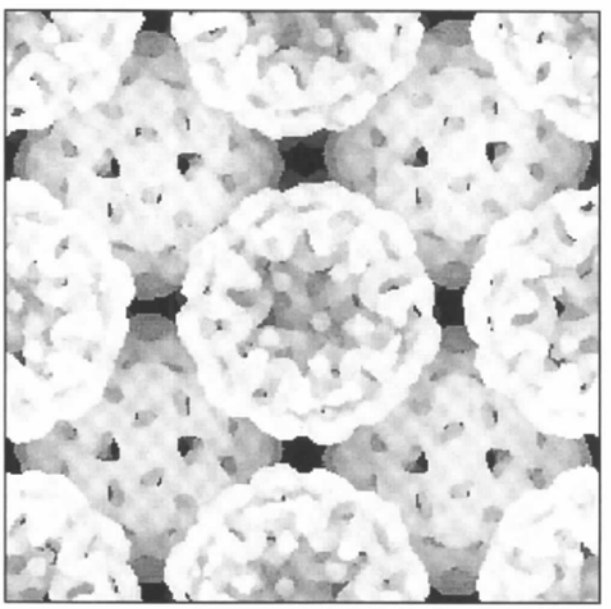


Fig. 5. A surface representation of the EM reconstruction into the unit cell illustrating the packing arrangement of particles in the cubic unit cell. This is the model that was used for the initiation of phase determination.

By inspection of an electron-density map calculated at 6.5 Å resolution, it became clear that the phases corresponded to the Babinet opposite solution, giving rise to a map in which the density is inverted ($\rho' = -\rho$). The inverted map showed that there were four copies of the D scaffolding protein per icosahedral asymmetric unit, not two as had been expected based on the EM reconstruction (Ilag *et al.*, 1995). It was also obvious that the D protein was highly helical in character, and it was possible to trace the chain of each of the D proteins. Comparison with the known structures of the F and G proteins from the virion also established the correct hand of the map, which was found to be opposite to that of the EM reconstruction.

To obtain a higher resolution map, however, it was necessary to turn to the warm data. The orientations and positions had to be re-determined, as these were significantly different from those in the frozen crystals due to shrinkage of the cell upon freezing. A coarse grid search using the new 6.5 Å resolution model was carried out as described above. After this, the climb procedure, also described above, was carried out at 6.5 Å resolution for both particles, until convergence had been reached. Changes of up to 1.4° in orientation and 1.8 Å in position were found relative to the frozen data (Table 2). As the data was rather incomplete, especially beyond 4.5 Å resolution, scaled F_{calc} amplitudes were used for reflections for which there were no F_{obs} measurements. This ensured that the map did not deteriorate due to the incompleteness of the data. During most of the averaging, 20-fold NCS averaging of each particle separately was used. Full 40-fold averaging was only introduced in the final stages of averaging, but did not lead to a significant improvement of the map. A number of different weighting schemes were tested (Arnold *et al.*, 1987), but these also did not appear to make any significant difference to the averaged map. The resolu-

Table 3. *Final correlation coefficient and R factors as a function of resolution*

Resolution range (Å)	CC†	R†
∞–15.70	0.924	17.5
15.70–11.15	0.939	15.1
11.15–9.12	0.939	15.5
9.12–7.90	0.933	16.2
7.90–7.07	0.924	17.4
7.07–6.46	0.912	18.8
6.46–5.98	0.896	20.3
5.98–5.59	0.888	21.2
5.59–5.27	0.889	21.4
5.27–5.00	0.882	22.1
5.00–4.77	0.860	24.1
4.77–4.57	0.820	26.9
4.57–4.39	0.793	28.7
4.39–4.23	0.766	30.6
4.23–4.09	0.741	31.7
4.09–3.96	0.669	36.3
3.96–3.84	0.576	39.8
3.84–3.73	0.587	39.9
3.73–3.63	0.611	39.4
3.63–3.54	0.631	37.9
Overall	0.868	23.3

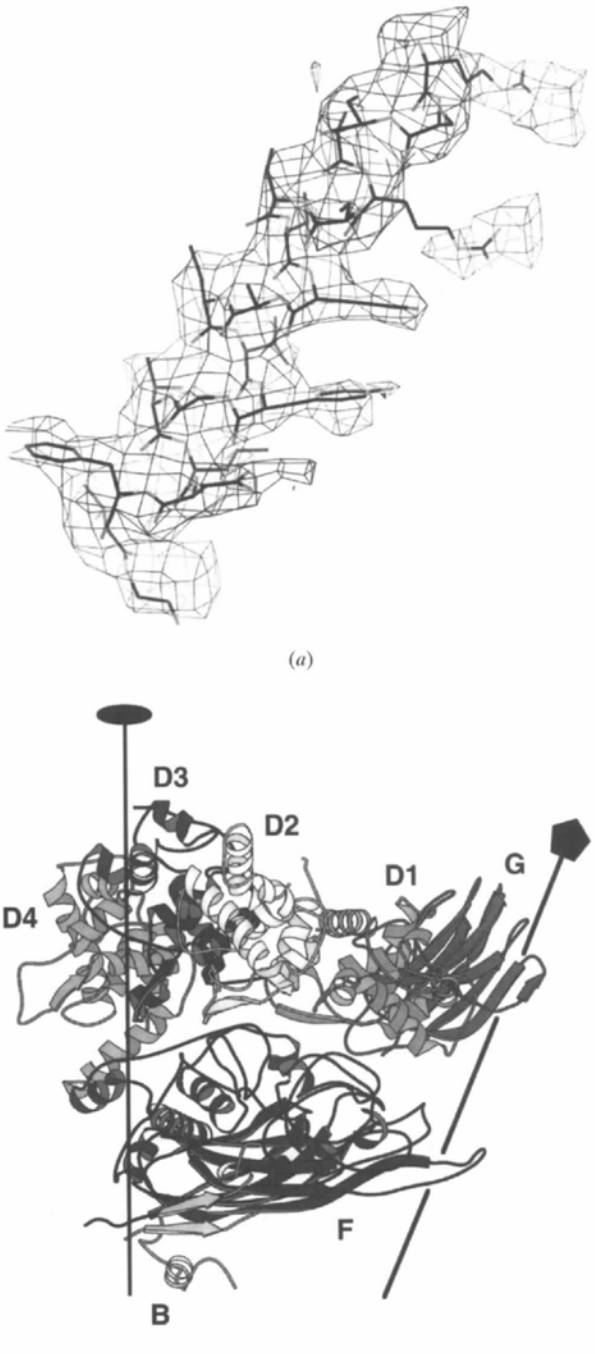
† Correlation coefficient (CC) and *R* factor (*R*) are defined in Table 2.

tion was extended two reciprocal lattice units at a time until the 3.5 Å resolution limit of the data had been reached. The final correlation coefficient and *R* factors as a function of resolution are shown in Table 3.

In spite of the low completeness of the data at higher resolution, the known amino-acid sequences of most of the D, F and G proteins and part of the B protein could be unambiguously built into the 3.5 Å electron-density map (Fig. 6). This was the first near-atomic resolution structure of a viral scaffolding protein and provided extensive insight into the assembly pathway mechanism (Dokland *et al.*, 1997).

8. Crystallographic refinement

The coordinates of the atoms in the icosahedral asymmetric unit were refined by conjugate-gradient minimization using the program *X-PLOR*, version 3.1 (Brünger, 1992). Seven independent chains were refined. These were the four independent D proteins (152 residues each), the F protein (426 residues), the G protein (175 residues) and the amino-terminal part (eight residues) and the carboxy-terminal part (41 residues) of the B protein. Only the data between 8.0 and 3.5 Å resolution was used, and strict NCS with a total of 40 constraints (20 for each independent particle) was imposed throughout the refinement. The X-ray crystallographic energy term minimized was $W_a/N_a \times \sum_h w_h(|F_{\text{obs},h}| - k|F_{\text{calc},h}|)^2$, where W_a is the overall weight between the X-ray crystallographic and the geometric energy terms, N_a is a resolution-dependent



(*a*)

(*b*)

Fig. 6. (*a*) Detail of the electron density from the final map at 3.5 Å resolution, showing part of helix 3 in the D1 scaffolding protein subunit. (*b*) Ribbon diagram showing the four copies of the D scaffolding protein (D1–D4), the underlying F capsid protein, the G spike protein and part of the internal scaffolding protein B in the icosahedral asymmetric unit. A twofold and a fivefold axis of symmetry are indicated.

Table 4. *Refinement statistics*

Refinement step	$R_{working}$†
Initial model	42.1
40 cycles REPEL (no electrostatic potential)	38.9
40 cycles Powell	35.9
Regeneration of model and recalculation of W_a	
50 cycles Powell	33.6
20 cycles Powell	33.2
Calculation of overall B factor	
20 cycles Powell	31.6

† $R_{working} = 100 \times \sum_h |F_{calc,h} - kF_{obs,h}| / \sum_h |F_{calc,h}|$, where $k = \langle F_{calc} \rangle / \langle F_{obs} \rangle$ for all reflections within each resolution bin.

normalization factor, h are the Miller indices of the structure factors, w_h are the weights for the individual reflections and k is a scale factor. The overall weight W_a was determined by running a short molecular dynamics calculation and comparing the normal of the gradient of the empirical energy term to the X-ray energy term (Brünger, 1992). The initial R factor was 42.1%. After 170 cycles of Powell minimization, the first 40 with the electrostatic potential terms switched off ('repel' minimization), the R factor had improved to 31.6% with reasonable geometry (Table 4). No R_{free} was calculated, as the high non-crystallographic redundancy causes R_{free} and $R_{working}$ to be almost identical (Hadfield *et al.*, 1997). Individual B-factor refinement was not carried out due to the limited resolution of the data. About 1.6% of the amino acids in the final structure were outside the generously allowed regions of the Ramachandran diagram, as expected from the limited resolution. The refined coordinates have been deposited with the Brookhaven Protein Data Bank.†

9. Molecular replacement with a poor model

Various starting models have been used to solve icosahedral virus structures, including X-ray structures of related or unrelated viruses, EM reconstructions and spherical shell models. For example, in the structure determination of bacteriophage MS2, Valegård *et al.* (1990) used the structure of the unrelated plant virus southern bean mosaic virus (Abad-Zapatero *et al.*, 1980) to initiate the phasing. Molecular replacement yielded a map which was not directly interpretable, but which was sufficiently good to define the positions of the heavy atoms in isomorphous heavy-atom derivatives and also showed that the map represented the Babinet opposite solution. Once accurate particle positions had been defined by using heavy-atom sites, averaging yielded an interpretable map which was completely different from the starting model (Valegård *et al.*, 1990, 1994). Similarly, the φX174 virion structure determination was initiated using the unrelated cowpea mosaic virus (Chen *et al.*, 1990) as a starting model (McKenna, Xia, Willingmann, Ilag, Krishnaswamy *et al.*, 1992; McKenna, Xia, Willingmann, Ilag & Rossmann, 1992). Also in this case, the phases had refined to the Babinet opposite solution, but did yield an interpretable map once the position of the particle in the unit cell had been determined accurately. Again, it was found that the starting model had little relationship to the final structure. The only relationship between the model and the final structure in these two cases was that they had similarly sized and shaped icosahedral envelopes.

In the present work, a starting model was obtained by fitting the atomic structure of the mature virion into the procapsid EM reconstruction and using an arbitrarily chosen protein structure for the unknown parts (Ilag *et al.*, 1995). In retrospect, it is clear that this starting model was very poor, since the hand of the EM reconstruction was wrong, the F capsid protein was rotated by about 90° from the correct orientation, only 120 copies – rather than 240 – of the D protein were modeled, and the D protein was in fact almost entirely α-helical while the G protein used as a model to represent the D proteins has a predominantly β structure. Furthermore, the EM reconstruction represents the open procapsid structure, while the X-ray data corresponded to the closed procapsid, in which the structures of the F and G proteins are similar to those in the virion (Dokland *et al.*, 1997). In hindsight, it would probably have been better to have used the virion X-ray structure directly, possibly after adding some density for the D protein. Nevertheless, placing this starting model into the crystal cell in the positions and orientations given by the Patterson and rotation functions, it was possible to find a resolution range (20–13 Å) which gave a reasonable correlation and R factor between F_{calc} and F_{obs}. A comparison of the phases between the 16-particle starting model (cycle B1) and the final averaged frozen data set (cycle B447) shows little similarity (Fig. 7*a*). Nevertheless, some agreement occurs for centric phases when $\alpha' = \alpha = 270°$ (the projected centres of symmetry are displaced by $\frac{1}{4}$ from the origin), which may be the reason for the eventual successful structure determination. Surprisingly, the phases for the first model (cycle A1), which contained only eight particles, are fairly well correlated with those of the second model (cycle B1), which contained 16 particles (Fig. 7*b*). In contrast, the phase difference between either starting model and the final solution is almost completely random (~90°). As mentioned above, at cycle B260, a Babinet inversion was found in the phases between 20 and 15 Å resolution. This is represented in the phase comparison plot between cycle B260 and B447 by the lines for which $\alpha' = \alpha + \pi$ (Fig. 7*c*). Thus, the two opposite solutions were 'fighting' each other near 15 Å resolution. After removing the data between 20 and 14.5 Å resolution, the

† Atomic coordinates and structure factors have been deposited with the Protein Data Bank, Brookhaven National Laboratory (Reference: 1ALO).

phases were extended in both directions until the 6.5 Å resolution limit of the frozen data had been reached. This data set was used successfully as a starting model for the phasing of the warm data. The final phases of the frozen data set (cycle B447) show relatively small differences compared to the phases of the final warm data set (cycle C450; Fig. 7*d*). Although the atoms in the starting model had the correct handedness, the overall features, which were based on the EM reconstruction, had the incorrect hand. At the low resolution (20–13 Å) at which the molecular replacement was initiated, it might have been expected that the overall features would be more important in defining the hand of the final structure, yet the structure after phase improvement and extension to 3.5 Å resolution was found to have the correct hand.

Ideally, an EM reconstruction, which often can be obtained relatively easily, would be sufficient in itself to provide a phasing start. The problem lies in the lack of resolution range overlap between the EM and X-ray data, although with higher resolution EM reconstructions becoming available (Böttcher *et al.*, 1997), this problem will be reduced. One case where the phasing was initiated directly from an EM reconstruction was for an insect parvovirus, *Galleria melonella* densovirus (Simpson *et al.*, 1998), for which the phasing was started using a resolution range between 50 and 25 Å. Alternatively, the EM reconstruction can be filled with atoms either on a regular grid or at pseudo-random positions to generate a starting model which extends to somewhat higher resolution than the EM reconstruction itself. In principle this approach is equivalent to starting the phasing with a structurally unrelated atomic model, and was used successfully in the structure determination of Norwalk virus (B. V. V. Prasad, personal communication). A logical extension of this would be to start off the phasing with a simple spherical shell model. The latter approach was used for canine parvovirus (Chapman *et al.*, 1992; Tsao, Chapman, Wu *et al.*, 1992; Tsao, Chapman & Rossmann, 1992), where it was also shown how the parameters for the spherical shell model could be determined directly from the data without any external information.

In this study, a number of starting models had to be abandoned because they did not converge to reasonably good *R* factors even after several cycles of averaging.

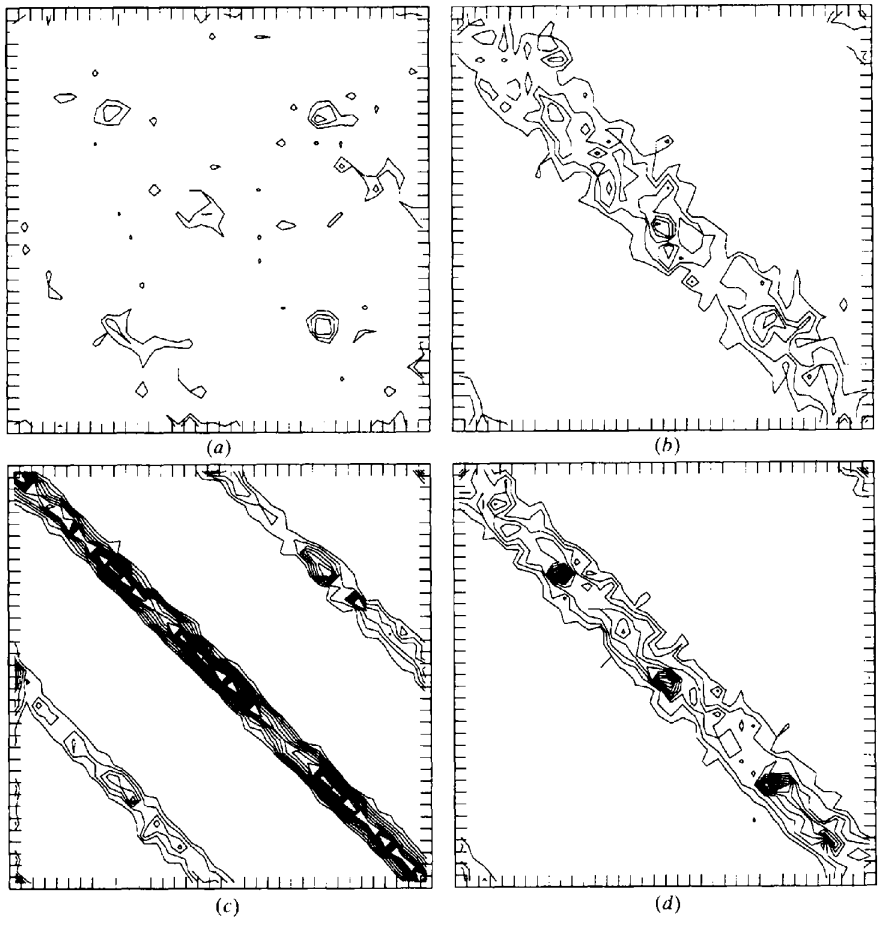


Fig. 7. Phase comparison plots, in which the phase of each reflection in one data set is plotted against the phase of the corresponding reflection in another data set. The contours represent the frequency of occurrence. All the comparisons are done in the resolution range from 20 to 13 Å. In (*a*) the phases of cycle B1 are compared to the final phases of the frozen data set (cycle B447). In (*b*) the phases for the eight-particle starting model (cycle A1) are compared with the phases for the 16-particle starting model (cycle B1). In (*c*) the phases of the last averaging cycle (B260) before the low-resolution Babinet opposite data was removed are compared to the phases of the final frozen data (cycle B447). In (*d*) the phases of the final frozen data (cycle B447) are compared with the phases of the final warm data (cycle C450).

One sensitive measure that was used to assess whether the starting model was successful was the frequency distribution of the electron density in the averaged maps. If the orientation and position of the particles in the starting model were correct, the averaged map had values only slightly lower than those of the unaveraged map, otherwise it was dominated by values close to zero.

Starting molecular replacement with poor models is possible due to the high non-crystallographic redundancy available in the icosahedral virus structures. In the present study, the crystals contained 40-fold NCS, yet only 20-fold symmetry was applied during most of the averaging without any noticeable detrimental effects. Arnold & Rossmann (1986) showed that the phasing power of the NCS averaging is proportional to the square root of the total redundancy (non-crystallographic redundancy times data completeness). As the averaging power is reduced, however, it becomes increasingly more important to have an accurate molecular envelope (Acharya *et al.*, 1989). In this study, a simple hollow shell was used for the envelope throughout. For simian virus 40, where likewise only fivefold symmetry was available, the map improved significantly when the density was averaged over the quasi-fivefold pentamers (Liddington *et al.*, 1991). In the present study, scaled F_{calc} amplitudes were used where there were no F_{obs} measurements. This has the effect of improving the data completeness.

Accurate knowledge of particle orientation and position may be especially important when the model is poor. In some fortunate cases, the particles are situated on special positions in the unit cell, thus allowing the orientation and position to be known exactly. However, this may cause problems when the starting model is a hollow shell or a low-resolution model, which has an at least approximate centre of symmetry. When the particle is sitting on a special position, the initial model-based phases will be centric. If the NCS elements have a centrosymmetric disposition in the space group, it would be impossible to break away from the centric starting phases of the model by density averaging. A way of solving this problem would be through the introduction of some asymmetry in the initial model, either by using a more realistic model or by arbitrarily introducing non-centric phases to some large low-resolution reflections. More often, the particles found in virus crystals exhibit slight deviations from being situated at special positions (Muckelbauer *et al.*, 1995; Zlotnick *et al.*, 1997), such as the 6.7 Å shift and less than 2° rotation seen in the present structure (Table 2). In this case, the initial phases will not be centric even with a spherical starting model, but there are likely to be considerable difficulties in determining the precise positional and orientational parameters. The search for a suitable starting model and finding the correct orientational parameters took considerable effort and time in the initial stages of the structure determination described here. The lessons learned here will hopefully permit faster and smoother structure determinations in the future.

Infected cells containing φX174 procapsid were provided through our long-standing collaborations with Nino Incardona and Ben Fane, who also contributed with invaluable discussions and support. We are indebted to Vic Ilag, who first isolated the procapsid in our laboratory and obtained procapsid crystals suitable for X-ray analysis. We would also like to thank the staff at CHESS and Robert Sweet at Brookhaven beamline X12c for providing excellent facilities and user support. TD was supported through a European Molecular Biology Organization (EMBO) postdoctoral fellowship for part of the time involved in this work. MGR was supported by an NSF grant. Additional support was provided through a Lucille Markey Foundation grant and a Purdue University re-investment grant.

References

Abad-Zapatero, C., Abdel-Meguid, S. S., Johnson, J. E., Leslie, A. G. W., Rayment, I., Rossmann, M. G., Suck, D. & Tsukihara, T. (1980). *Nature (London)*, **286**, 33–39.

Acharya, R., Fry, E., Stuart, D., Fox, G., Rowlands, D. & Brown, F. (1989). *Nature (London)*, **337**, 709–716.

Arnold, E. & Rossmann, M. G. (1986). *Proc. Natl Acad. Sci. USA*, **83**, 5489–5493.

Arnold, E., Vriend, G., Luo, M., Griffith, J. P., Kamer, G., Erickson, J. W., Johnson, J. E. & Rossmann, M. G. (1987). *Acta Cryst.* A**43**, 346–361.

Böttcher, B., Wynne, S. A. & Crowther, R. A. (1997). *Nature (London)*, **386**, 88–91.

Brünger, A. T. (1992). *X-PLOR, Version* 3.1. *A System for X-ray Crystallography and NMR*. New Haven and London: Yale University Press.

Burgess, A. B. (1969). *Proc. Natl Acad. Sci. USA*, **64**, 613–617.

Chapman, M. S., Tsao, J. & Rossmann, M. G. (1992). *Acta Cryst.* A**48**, 301–312.

Chen, Z., Stauffacher, C. V. & Johnson, J. E. (1990). *Sem. Virol.* **1**, 453–466.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.

Dokland, T., McKenna, R., Ilag, L. L., Bowman, B. R., Incardona, N. L., Fane, B. A. & Rossmann, M. G. (1997). *Nature (London)*, **389**, 308–313.

Ekechukwu, M. A., Oberste, D. J. & Fane, B. A. (1995). *Genetics*, **140**, 1167–1174.

Evans, P. R. (1993). *Proceedings of the CCP4 Study Weekend, 29–30 January 1993*, edited by L. Sawyer, N. Isaacs & S. Bailey, pp. 114–123. Warrington: Daresbury Laboratory.

French, S. & Wilson, K. (1978). *Acta Cryst.* A**34**, 517–525.

Fujisawa, H. & Hayashi, M. (1977*a*). *Virology*, **24**, 303–313.

Fujisawa, H. & Hayashi, M. (1977*b*). *Virology*, **21**, 506–515.

Fujisawa, H. & Hayashi, M. (1977*c*). *J. Virol.* **23**, 439–442.

Hadfield, A. T., Lee, W.-M., Zhao, R., Oliveira, M., Minor, I., Rueckert, R. R. & Rossmann, M. G. (1997). *Structure*, **5**, 427–441.

Hall, C. E., Maclean, E. C. & Tessman, I. (1959). *J. Mol. Biol.* **1**, 192–194.

Hayashi, M., Aoyoma, A., Richardson, D., Jr. & Hayashi, M. (1988). *The Bacteriophages*, Vol. 2, edited by R. Calendar, pp. 1–71. New York: Plenum Press.

Ilag, L. L., Olson, N. H., Dokland, T., Music, C. L., Cheng, R. H., Bowen, Z., McKenna, R., Rossmann, M. G., Baker, T. S. & Incardona, N. L. (1995). *Structure*, **3**, 353–363.

Incardona, N. L., Tuech, J. K. & Murti, G. (1985). *Biochemistry*, **24**, 6439-6446.

Johnson, J. E., Rossmann, M. G., Smiley, I. E. & Wagner, M. A. (1974). *J. Ultrastruct. Res.* **46**, 441–451.

Liddington, R. C., Yan, Y., Moulai, J., Sahli, R., Benjamin, T. L. & Harrison, S. C. (1991). *Nature (London)*, **354**, 278–284.

McKenna, R., Bowman, B. R., Ilag, L. L., Rossmann, M. G. & Fane, B. A. (1996). *J. Mol. Biol.* **256**, 736–750.

McKenna, R., Ilag, L. L. & Rossmann, M. G. (1994). *J. Mol. Biol.* **237**, 517–543.

McKenna, R., Xia, D., Willingmann, P., Ilag, L. L., Krishnaswamy, S., Rossmann, M. G., Olson, N. H., Baker, T. S. & Incardona, N. L. (1992). *Nature (London)*, **355**, 137–143.

McKenna, R., Xia, D., Willingmann, P., Ilag, L. L. & Rossmann, M. G. (1992). *Acta Cryst.* B**48**, 499–511.

Muckelbauer, J. K., Kremer, M., Minor, I., Tong, L., Zlotnick, A., Johnson, J. E. & Rossmann, M. G. (1995). *Acta Cryst.* D**51**, 871–887.

Mukai, R., Hamatake, R. K. & Hayashi, M. (1979). *Proc. Natl Acad. Sci. USA*, **76**, 4877–4881.

Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.

Reynolds, E. S. (1963). *J. Cell Biol.* **17**, 208.

Rodgers, D. W. (1997). *Methods Enzymol.* **276**, 183–203.

Rossmann, M. G. (1990). *Acta Cryst.* A**46**, 73–82.

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

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.-b., Wu, H., Choi, H.-K. & Lynch, R. E. (1992). *J. Appl. Cryst.* **25**, 166–180.

Siden, E. J. & Hayashi, M. (1974). *J. Mol. Biol.* **89**, 1–16.

Simpson, A. A., Chipman, P. R., Baker, T. S., Tijssen, P. & Rossmann, M. G. (1998). *Structure*. In the press.

Sinsheimer, R. L. (1959). *J. Mol. Biol.* **1**, 37–42.

Tonegawa, S. & Hayashi, M. (1970). *J. Mol. Biol.* **48**, 219–242.

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

Tsao, J., Chapman, M. S. & Rossmann, M. G. (1992). *Acta Cryst.* A**48**, 293–301.

Tsao, J., Chapman, M. S., Wu, H., Agbandje, M., Keller, W. & Rossmann, M. G. (1992). *Acta Cryst.* B**48**, 75–88.

Valegård, K., Liljas, L., Fridborg, K. & Unge, T. (1990). *Nature (London)*, **345**, 36–41.

Valegård, K., Murray, J. B., Stockley, P. G., Stonehouse, N. J. & Liljas, L. (1994). *Nature (London)*, **371**, 623–626.

Weber, P. C. (1997). *Methods Enzymol.* **276**, 13–22.

Weisbeek, P. J. & Sinsheimer, R. L. (1974). *Proc. Natl Acad. Sci. USA*, **71**, 3054–3058.

Zlotnick, A., Natarajan, P., Munshi, S. & Johnson, J. E. (1997). *Acta Cryst.* D**53**, 738–746.