Acta Cryst. (1998). D**54**, 1295–1305

Structure Determination of *Nudaurelia capensis* ω Virus

Sanjeev Munshi,† Lars Liljas‡ and John E. Johnson*

Department of Molecular Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, USA. E-mail: jackj@scripps.edu

(*Received 22 August 1997; accepted 23 March 1998*)

Abstract

The structure of *Nudaurelia capensis* ω virus (NωV), a single-stranded RNA virus, was determined to 2.8 Å resolution. Triclinic crystals (a = 413.6, b = 410.2, c = 419.7 Å, α = 59.13, β = 58.9, γ = 64.0°) diffracted X-rays beyond 2.7 Å resolution. The unit cell contained one icosahedral virus particle, providing 60-fold non-crystallographic symmetry (n.c.s.) and structural redundancy. The particle orientation in the unit cell was determined by self-rotation function analyses. Initial phases to 18 Å resolution were derived from a hollow spherical model of 192 Å outer radius and 139 Å inner radius, filled with uniform electron density. Radii of the model were determined by maximizing the correlation of the model-based calculated data with the low-resolution X-ray diffraction and solution-scattering data. Phases were refined by 60-fold non-crystallographic electron-density averaging and extended in small steps to a resolution of 5 Å. The phases obtained represented a mixture of four different phase sets, each consistent with the icosahedral symmetry constraints. The resulting electron density was not interpretable. A difference Fourier map computed with the native and an isomorphous heavy-atom derivative data sets and phases refined by real-space averaging was interpretable only if data within the 10 Å resolution shell were used. Maps calculated with data significantly higher than 10 Å resolution failed to display a constellation of heavy-atom sites consistent with the $T = 4$ icosahedral symmetry. Attempts to extend the phases beyond 10 Å resolution, starting with either phases based on a model or single isomorphous replacement, were unsuccessful. Successful phase extension was achieved by computing the phases for the higher resolution reflections from a partial atomic model (poly gly) built into the averaged 10 Å electron-density map. Phases from this model served as the starting point for n.c.s. phase refinement and extension to slightly higher resolution. The atomic model was improved at each extension interval and these phases were used for the subsequent phase calculation and extension. The entire polypeptide backbone corresponding to the NωV structure was built into the map at 4 Å. The same procedure for phase refinement was used to extend the phases to 2.8 Å in small increments of resolution. The overall molecular averaging R factor and correlation coefficient at 2.8 Å resolution were 18.4% and 0.87, respectively.

1. Introduction

Use of non-crystallographic symmetry as a tool for phase refinement and the determination of phases at incrementally higher resolution (phase extension) has led to the atomic resolution models of several oligomeric proteins. The methodology used for the phase refinement and extension by density averaging has been extensively reviewed (Rossmann, 1972, 1990; Jones, 1992). Electron-density maps of proteins with non-crystallographic redundancy computed with phases derived from heavy-atom derivatives are often noisy and difficult to interpret. Most of these maps are interpretable following the averaging of the parts of the structure related by non-crystallographic symmetry. The power of density averaging has been extensively exploited in the structure determination of the protein capsids of several icosahedral viruses (Grant *et al.*, 1992; Liddington *et al.*, 1991; McKenna *et al.*, 1992; Speir *et al.*, 1995; Yan *et al.*, 1996). Initial phases have often been derived either from heavy-atom derivatives (Harrison *et al.*, 1978; Larson *et al.*, 1993; Liddington *et al.*, 1991; Liljas *et al.*, 1982; Rossmann *et al.*, 1985) or, at low resolution, from the homologous atomic models (Acharya *et al.*, 1989; Agbandje *et al.*, 1993; Chen *et al.*, 1989; Golmohammadi *et al.*, 1996; Luo *et al.*, 1987, 1992). Using the appropriate initial phasing model, phase refinement and extension to high resolution has been carried out for several icosahedral viruses. Successful attempts of *ab initio* phasing at low resolution have been reported for the structure determination of southern bean mosaic virus (Johnson *et al.*, 1976) and polyoma virus (Rayment, 1983; Rayment *et al.*, 1982). Starting from a hollow-sphere model, *ab initio* phasing at low resolution followed by successful extension to 8 Å resolution has been reported for the structure determi-

† Current address: WP44-B122, Merck & Co. Inc., West Point, Pennsylvania 19486, USA.

‡ Current address: Department of Molecular Biology, Uppsala University, Uppsala Biomedical Centre, Box 590, S-751 24 Uppsala, Sweden.

ISSN 0907-4449

nation of canine parvovirus (Tsao *et al.*, 1992). The structure had previously been determined with SIR phases to 3.2 Å resolution. Despite the success of density averaging as a tool for phase refinement and extension in the structure determination of icosahedral viruses, difficulties and failures in the structure determination of structurally related viruses have also been reported (Acharya *et al.*, 1989; Ban & McPherson, 1995; Larson *et al.*, 1993). Here we report the structure determination of NωV at 2.8 Å resolution by using a hollow spherical shell as the initial phasing model. Initial phases were computed to 18 Å resolution and refined by real-space density averaging using 60-fold non-crystallographic symmetry. The phases were extended to high resolution in small intervals of reciprocal space by Fourier back-transformation of the averaged electron-density map. However, this phase-extension approach was unsuccessful, owing to the accumulation of phases for different reflections that satisfied the icosahedral symmetry and corresponded to related, but different, electron-density distributions. An alternative approach to phase extension was required to determine the structure at high resolution. The novel features of the problems and their solutions are presented.

NωV is a spherical positive-sense single-stranded RNA virus that infects the South African pine emperor moth (Hendry *et al.*, 1985). The icosahedral shells are 410 Å in diameter and consist of 240 copies of chemically identical subunits arranged in a $T = 4$ quasi-symmetric lattice (Johnson *et al.*, 1994; Munshi *et al.*, 1996). Encapsidated in the shell are two RNA molecules, RNA1 (5000 bp) and RNA2 (2500 bp) (Agrawal & Johnson, 1995). RNA2 encodes the capsid protein precursor α [70 kDa: 644 amino acids (Agrawal & Johnson, 1992)]. The 70 kDa protein α exists as two polypeptides β (62 kDa) and γ (8 kDa). The bond between Asn570 and Phe571 is cleaved by an assembly-dependent autoproteolytic mechanism (Agrawal & Johnson, 1995). Several features of the genetic organization and capsid assembly of NωV are similar to the members of $T = 3$ nodaviruses (Johnson *et al.*, 1994). However, these two different classes of viruses display no significant amino-acid sequence homology in their capsid proteins (Agrawal & Johnson, 1992).

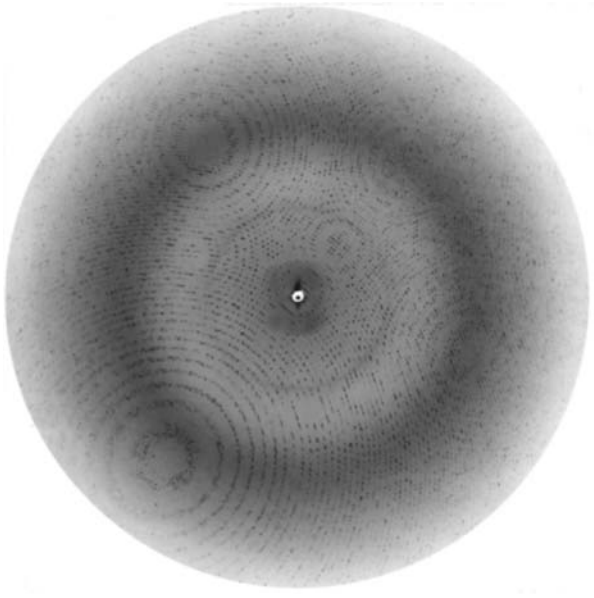


Fig. 1. A 0.3° oscillation photograph from a crystal of NωV recorded on the A1 station at the Cornell High Energy Synchrotron Source. The wavelength used was 1.565 Å and the exposure time was 90 s. The crystal-to-film distance was 100 mm. Data were measured to 2.8 Å resolution. A total of 4935 full and 20710 partial reflections are predicted to be on the photograph. 3704 full and 12024 partial reflections were measured with $I/\sigma(I) > 4.0$.

2. Crystallization

The purification of NωV has been described previously (Hendry *et al.*, 1985). NωV was crystallized using the sitting-drop method of vapor diffusion. Crystals were grown as reported earlier (Cavarelli *et al.*, 1991) and belong to space group $P1$ with $a = 413.6$, $b = 410.2$, $c = 419.7$ Å, $\alpha = 59.13$, $\beta = 58.9$, $\gamma = 64.0°$. The unit cell contains one complete virus particle, resulting in 60-fold non-crystallographic redundancy in the capsid structure. The crystals diffract beyond 2.7 Å resolution.

3. Data collection and processing

Initial data were collected on photographic film on beamline X-12C at the National Synchrotron Light Source (NSLS). Additional data were collected at Cornell High Energy Synchrotron Source (CHESS), with photographic film on the A1 and F1 stations (Fig. 1). The strategy used for data collection and processing was similar to that reported earlier (Cavarelli *et al.*, 1991). During data collection crystals were not aligned in any specific orientation. Depending on the size and the quality of the crystals, 1–5 exposures were recorded from each crystal. More than 200 crystals were used to record data on 850 films. A pack of two films (*A* and *B*) was used for each 0.3° oscillation picture recorded. The amplitudes of the lower order reflections that were underestimated on the *A* film, owing to the saturation limits of the photographic film, were better estimated by processing the data from the *B* films. The data were processed and post-refined with the *Purdue* film-processing package (Rossmann, 1979; Rossmann *et al.*, 1979). Reflections with a partiality estimate greater than 0.5, after appropriate scaling, were included in the data set. The final native data set consisted of 500 *A* films and 192 *B* films (Table 1).

In addition to the native data set, small amounts of data were collected for four potential heavy-atom derivatives (Table 2). Four different compounds, potassium hexachloroplatinate (PHCP), fluorescein mercury

Table 1. *Data collection for native triclinic crystals of NωV*

Number of films	500 (includes 192 pairs of *A*/*B* scaled films)
Processing resolution limit (Å)	2.8
$I/\sigma(I)$ cut off	2 (in processing films)
Total number of observations	4141071
Number of independent observations	2465842
R_{merge} (%)	12.1
Data completeness	
Resolution (Å)	Theoretically possible data (%)
∞–20	82
20–15	83
15–10	83
10–8	82
8–6	77
6–5	76
5–4	73
4–3	49

acetate (FMA), potassium tetrachloroaurate (PTCA) and trimethyl lead acetate (TMLA) were used. About 5–7% of data were collected from the crystals soaked in 2–5 m*M* solutions of heavy-atom compounds for 8–21 h. Heavy-atom derivative data were processed to 4 Å resolution. Each heavy-atom data set was individually processed and scaled to the native data (Fig. 2).

4. Rotation function

The self-rotation function (Rossmann & Blow, 1962) was computed for the native and derivative data sets, in different resolution ranges. These results were nearly identical to those reported earlier (Cavarelli *et al.*, 1991).

5. Low-resolution models

Low-resolution phases were calculated with two types of models. The first model (referred to as model M192 in the following text) was a hollow spherical shell. The electron density was set to a uniform arbitrary value within the spherical shell and to zero outside the shell. The inner and outer radii of the shell were systematically optimized by an *R*-factor search between the low-resolution observed amplitudes and the amplitudes calculated from the spherical-shell model. The model with an inner radius of 139 Å and an outer radius of 192 Å resulted in the lowest *R* factor (54%) for data between 30 and 18 Å resolution. The value for the outer radius of the shell was consistent with the mean spherically averaged radius of the virus particle computed from the solution-scattering data (Johnson *et al.*, 1994) and also agreed with inter-particle contacts based on crystal packing. The second model was an atomic model based on the low-resolution electron-microscopy structure of NβV (Olson *et al.*, 1990). The model was constructed by transforming the known $T = 3$ capsid structure of black-beetle virus (BBV) into a $T = 4$ capsid structure. The 180 subunits of the $T = 3$ capsid of BBV were radially expanded to accommodate 60 more subunits of BBV to obtain the $T = 4$ model.

6. Low-resolution phase extension

Structure factors were computed for both models between 30 and 18 Å resolution and the calculated phases were used with the observed structure-factor amplitudes to calculate an electron-density map. This map was averaged using 60-fold non-crystallographic symmetry employing a spherical envelope defined by an inner radius of 100 Å and an outer radius of 205 Å. Tangential planes were defined to prevent the overlap of density associated with adjacent virus particles. The averaged map was Fourier transformed to yield a set of phases that were consistent with imposed icosahedral symmetry. The phases were refined by density averaging in a cyclic fashion until convergence (Rossmann, 1990). The procedure was performed on a parallel processing computer (Cornea-Hasegan *et al.*, 1995). 15 cycles of density averaging were performed with data between 30 and 18 Å resolution using weighted observed amplitudes. Phases were computed to 15 Å resolution by increasing the resolution in steps of one reciprocal lattice unit. Usually six cycles of phase refinement were performed after each extension. Despite high correlation coefficients and low *R* factors between the observed amplitudes and the amplitudes computed by Fourier transformation of the averaged electron-density maps, the maps were difficult to interpret. A 28 Å resolution electron-microscopy image reconstructed from the frozen hydrated samples of NωV (Johnson *et al.*, 1994) was used as a reference to judge the quality of the X-ray electron-density map. The quasi-symmetric arrangement of the subunits, as indicated by the electron-microscopy structure, was poorly represented in the X-ray electron-density map.

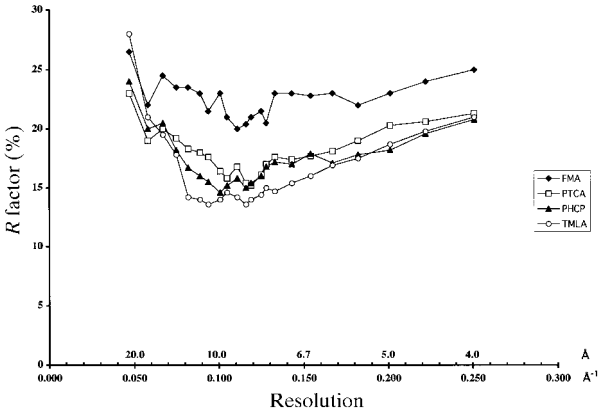


Fig. 2. *R* factors indicating the difference between the heavy-atom derivative and the native amplitudes as a function of resolution. $R = 100\left(\sum \left||F_N| - k|F_{HN}|\right| / \sum |F_N|\right)$, where F_N and F_{HN} are native and heavy-atom derivative structure amplitudes, respectively, and k is a scale factor.

Table 2. *Heavy-atom data collection and refinement*

Data collection

	FMA	Pt	Au	Pb
Number of films	21	18	19	16
Processing resolution limit (Å)	4	4	4	4
Independent observations (%)	7	5	6	6

FMA-derivative refinement analysis

Subunit	Fractional coordinates			Relative to icosahedral axes (Å)				Relative occupancy
	x	*y*	*z*	*X*	*Y*	*Z*	*D*	
A	−0.4075	0.2155	0.3976	99	2	164	191.5	10.2
B	−0.4156	0.0317	0.4114	38	−38	166	174.5	8.2
C	−0.4545	0.0749	0.3720	34	36	164	171.3	10.0
D	−0.4611	0.0588	0.3515	24	−36.5	166	171.6	9.7

Resolution range (Å)	30–4.0
Number of unique reflections phased	19979
Mean figure of merit	0.3

The success of *ab initio* phasing at low resolution critically depends on the accuracy and the completeness of the low-order diffraction data (Chapman *et al.*, 1992; Rayment, 1983). About 5% of the lower order amplitudes were underestimated to a varying extents in the Nω V data set owing to the saturation of the *A* films. However, lack of proper estimates for these lower order amplitudes was not initially thought to be critical for phase refinement. To rule out that possibility, amplitudes for most of these overloaded measurements were replaced by those from the *B* films, after film-to-film scaling. 192 *A*/*B* film pairs were scaled together and merged with the rest of the data. *Ab initio* phasing with the spherical-shell model M192 was attempted again using the improved native data set. The phases were refined at 18 Å resolution and extended to 10 Å resolution (Fig. 3). A spherical envelope with 210 Å outer radius and 100 Å inner radius was used for averaging. The electron-density map at 10 Å resolution clearly displayed the features expected of Nω V based on the electron-microscopy image (Figs. 4*a* and 4*b*). A molecular envelope was defined from the X-ray electron-density map at 10 Å resolution and subsequently used during density averaging. Phases were further refined and extended to 3.6 Å resolution in intervals of one reciprocal lattice unit, followed by a number of cycles of averaging until convergence was achieved. The quality of the electron-density maps steadily degraded as the resolution was increased beyond 10 Å. Discontinuity in the electron density and the poor definition of the quasi-symmetry at 6 Å resolution are shown in Fig. 5(*a*). The correlation coefficient and *R* factor remained reasonable (Fig. 5*b*) even when electron density was poor. The inability to extend the phases beyond 10 Å resolution was ascribed to the poor quality of phases at and beyond 10 Å.

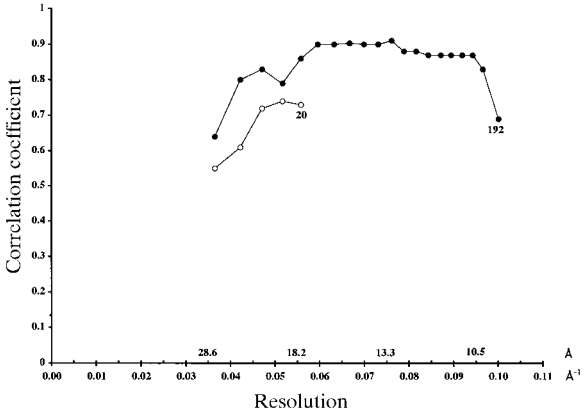
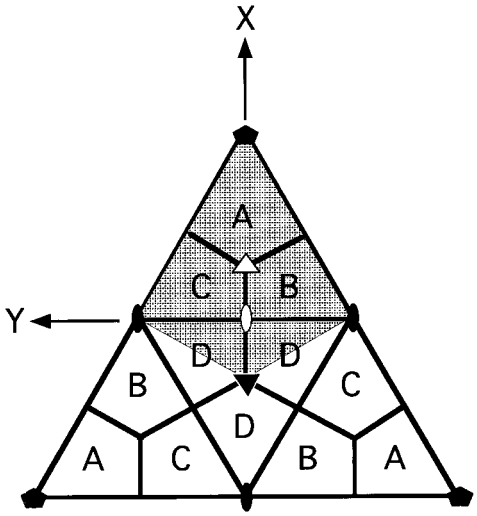


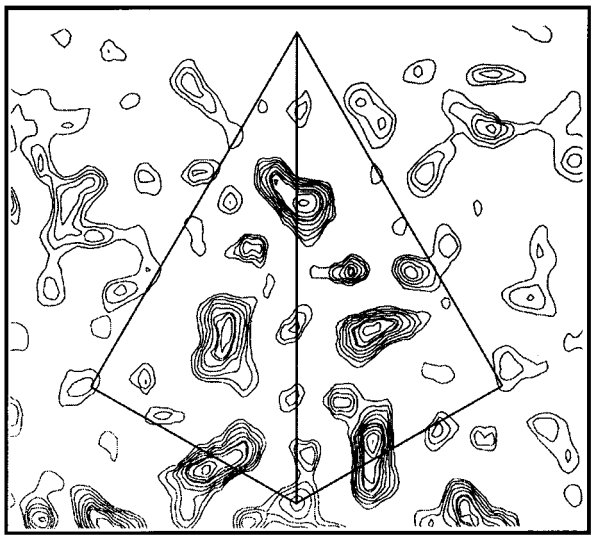
Fig. 3. Plot of correlation coefficient as a function of resolution. The initial phases between 30 and 18 Å resolution, computed from the hollow spherical shell model M192, were refined and extended to 10 Å resolution. The mean correlation coefficient and *R* factor at 18 Å resolution for 11802 independent reflections were 0.69 and 31%, respectively. At 10 Å resolution the mean correlation coefficient and *R* factor for 88854 independent reflections were 0.85 and 19.6%, respectively. The profiles are labeled with the number of cycles of averaging.

7. Difference Fourier analysis

The refined phases for data between 30 and 10 Å resolution were used to compute difference Fourier maps between the native data and the data from four different heavy-atom derivatives. The difference maps were averaged over the 60-fold non-crystallographic symmetry and the maps corresponding to the standard icosahedral asymmetric unit were plotted with respect to an orthogonal set of axes described in Fig. 4(*a*). The difference maps were featureless for platinum, gold and lead derivatives. However, peaks due to the heavy atoms related by the quasi-symmetry in the $T = 4$ capsid were observed in the difference maps with the FMA data. The peaks were observed as large negative values in the difference electron-density map (Fig. 6). This indicated



(*a*)



(*b*)

Fig. 4. (*a*) Diagrammatic representation of the subunit association on a face of a $T = 4$ icosahedron. Labeled trapezoids represent the individual capsid subunits. Asymmetric unit of the icosahedron contains four subunits *A*, *B*, *C* and *D* (shaded area). Subunits *A*, *B* and *C* are related by a quasi-threefold axis of symmetry (open triangle) and the subunits *C* and *D* are related by a quasi-twofold axis of symmetry (open oval). The two types of trimers *ABC* and *DDD* are related by a quasi-twofold axis of symmetry. *X* and *Y* define the right-handed orthogonal coordinate system used for displaying the icosahedral asymmetric unit. (*b*) Section of the electron-density map of NωV computed with the refined phases at 10 Å resolution. The starting low-resolution phases at 18 Å, computed from the hollow spherical shell model M192 were refined by density averaging and extended to 10 Å resolution. The area bounded by the quadrangle represents the icosahedral asymmetric unit. The density corresponds to a 2 Å slice at $Z = 164$ Å (from the center of the virus) along the icosahedral threefold axis. Distribution of the density corresponds to four subunits *A*, *B*, *C* and *D* of the icosahedral asymmetric unit. The relative orientation and the position of the density with respect to the icosahedral axes is consistent with the electron-microscopy image.

that the phases computed in the refinement process correspond to a negative image of the electron density – the Babinet-inverted image (Valegård *et al.*, 1991). The heavy-atom parameters, including the relative occupancy, temperature factor, position and scale factor, were refined (Table 2) using the heavy-atom least-squares refinement and phasing program with non-crystallographic constraints (Rossmann, 1976). The difference Fourier maps between native and FMA data were featureless if the phases (extended from model M192) significantly beyond 10 Å resolution were used. This analysis only validated the correctness of the refined phases to 10 Å resolution and strongly suggested significant problems beyond that resolution. The interpretability of the difference map appeared to be a valid criterion for evaluating the quality of the phases as different approaches to phase extension beyond 10 Å were attempted.

8. Phase extension from SIR phases

Phase refinement by density averaging and extension to high resolution were attempted starting with SIR phases computed between 20 and 8 Å resolution. The initial correlation coefficient was 0.22 and improved significantly to 0.85 upon density averaging. However, the electron-density maps lacked the features expected at 8 Å resolution. The SIR phases were combined with model M192 refined phases to 10 Å resolution and phase extension beyond 10 Å was attempted again. Despite reasonable correlation coefficients, the maps at higher resolution were uninterpretable.

9. Phase extension and refinement analysis based on model data

Persistent failure in successfully extending the phases beyond 10 Å resolution prompted the analysis of the phase-extension procedure using ideal calculated data. A model used to generate a test data set was constructed from the atomic coordinates (poly gly) of BBV and cowpea chlorotic mottle virus (CCMV) subunits. A subunit of BBV (44 kDa) and CCMV (19.8 kDa) were visually modeled into the electron density (for each of the four subunits of NωV) in the 10 Å map computed with refined model M192 phases. The BBV subunits fit well into the $T = 4$ shell region, while the CCMV subunits fit into the protruding density. The complete virus particle was generated by applying icosahedral symmetry and the atomic model was placed in the *P*1 cell of NωV in the proper orientation. A molecular envelope was generated from the atomic coordinates to be used for the density averaging. Structure factors based on the model coordinates were computed for data between 20 and 10.5 Å resolution (set I) and for data between 20 and 8 Å resolution (set II). Both phase sets were refined and extended to 7 Å resolution in

intervals of one reciprocal lattice unit using the calculated structure amplitudes to 7 Å resolution as the 'observed' data. In order to see if the problems with phase extension were caused by errors in the observed data or arose from an intrinsic problem with the procedure, the two extended and refined phase sets for data between 20 and 7 Å were compared with each other by examining the mean cosine phase difference as a function of resolution (Fig. 7*a*). A mean value of 1.0 would imply perfectly matching phases and a mean value of −1.0 would represent phases that are systematically shifted by 180° (Babinet solution). The selection of an enantiomer between α and $-\alpha$ or between $\pi - \alpha$ and $\pi + \alpha$ is, however, not reflected in the cosine plots. The two phase sets show little correlation beyond 8.5 Å resolution, indicating that the phase-extension procedure itself was breaking down, even with perfect data. Although the mean phase difference for data between 9.9 and 9.0 Å is only 29°, there are a significant number of reflections that have phases of opposite handedness (Fig. 7*b*). These results and other control experiments (data not shown) indicated a peculiar problem with the phase-extension process.

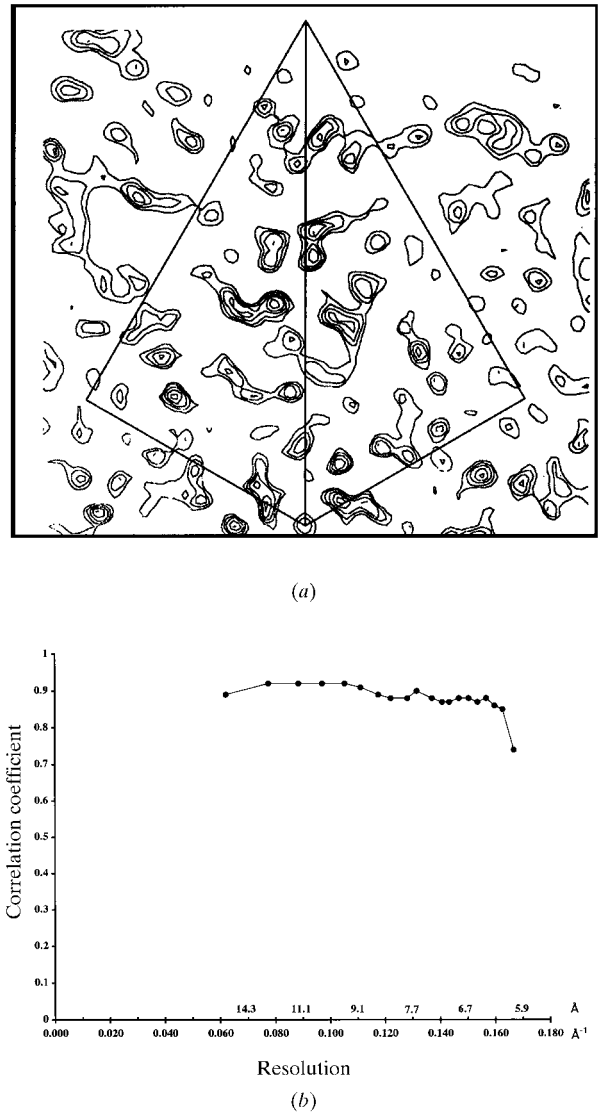


Fig. 5. (*a*) A 2 Å slice of the electron-density map at 6 Å resolution at Z = 164 Å along the icosahedral threefold axis. The map was computed with phases extended and refined to 6 Å, starting from the low-resolution phases computed from the spherical shell model M192. The phases were computed to 6 Å resolution by Fourier back-transformation of the averaged electron-density map. (*b*) Plot of correlation coefficient as a function of resolution for data between 20 and 6 Å resolution. The mean correlation coefficient and R factor for 379452 independent reflections were 0.89 and 17.6%, respectively.

10. Unconventional phase-extension procedure

The averaged electron-density map at a given resolution may contain errors due to partial convergence of the phases for reflections at the edge of the resolution range. Errors in the maps could also result from possible phase 'crossovers' (see below) carried over from the starting phase set. Hence, phases computed for reflections in the next shell of data in the extension by Fourier transformation of the map could result in incorrect phase assignments. Thus, for a reflection with α as the correct phase, $-\alpha$, $\pi - \alpha$ or $\pi + \alpha$ could be selected. Each of these four possibilities would satisfy the non-crystallographic-symmetry constraints and each defines an image related to, but clearly different from, that generated by α. The final map was not interpretable because the density is a composite of the correct image, the image with the phase changed in sign (the enantiomorph of the correct image) and the Babinet opposites of both these images. The density for all four of the images is perfectly consistent with icosahedral symmetry; thus, during the process of refinement in this pathological case, the phase for each reflection may converge to any one of the four possible phase sets.

The systematic analysis of the phase refinement and extension process with the perfect model-based data suggested that phases determined for the reflections in the shell at higher resolution during extension failed to maintain consistent handedness. During extension, the possible 'crossover' of phases from one set to the other could not be predicted *a priori*. After several cycles of density averaging, phases would cluster into sets of different handedness, resulting in an uninterpretable electron-density map.

In order to maintain consistent handedness of the phases for different reflections during extension and refinement, an alternative approach was attempted. The procedure was based on the assumption that the consistency in the handedness of the phases could be maintained if the phases for the extended reflections were computed from an atomic model fitted into the

existing electron-density map. The electron-density map at 10 Å resolution, computed from the refined model M192 phases, was used as a starting point. The well defined parts of the map were modeled (Jones *et al.*, 1991) with polypeptides (poly gly) from the known virus capsid structures of CCMV and BBV. Nearly the entire subunits of BBV and CCMV, appropriately trimmed, were visually fitted into the density interpreted as each subunit of NωV. The NωV capsid-protein subunit consisted of 644 amino acids and the initial model consisted of 470–490 amino acids for each of the four different subunits in the icosahedral asymmetric unit. Phases were computed from the atomic model to 9.5 Å resolution, combined with observed amplitudes and refined by six cycles of density averaging. Phases were extended to 9.1 Å resolution at an interval of half the reciprocal lattice unit. The averaged map was re-examined and the model was optimized into the map. Phases were next computed from the modified atomic model to a resolution 2–3 reciprocal lattice units higher than the resolution of the map, combined with observed amplitudes and refined by several cycles of density averaging. This approach was used in a cyclic fashion (Fig. 8).

To minimize model bias, several segments of electron density in the map were left unmodeled and the atomic model was built in a few regions of the map that had no defined density. During phase refinement, correct but unmodeled density was recovered and no density for the incorrectly modeled atoms was visible in the map. As the resolution was increased, the model was optimized and a molecular envelope with better definition was generated. The validity of the successful phase extension was confirmed by analyzing the difference Fourier maps between native and FMA data, using the newly refined phases beyond 10 Å resolution (Fig. 9).

Phases were extended and refined, using the procedure described above, to 4 Å resolution. The electron-density map at 4 Å resolution displayed well defined

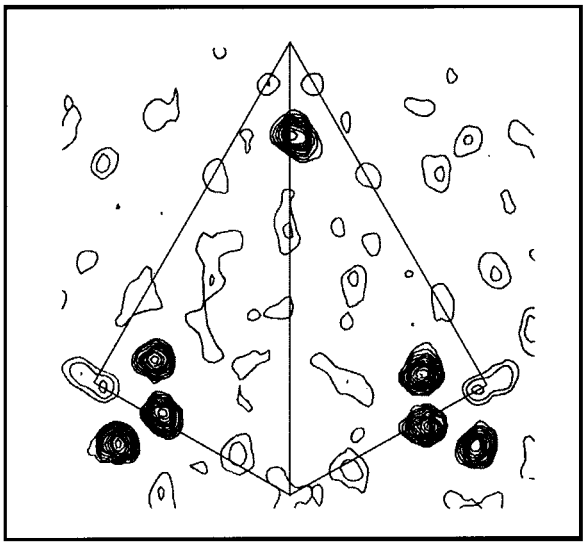


Fig. 6. The difference Fourier map for data between 30 and 10 Å resolution represents a 2 Å slice at Z = 164 Å along the icosahedral threefold axis. The map was computed with phases from the hollow spherical-shell model M192, refined to 10 Å resolution and contoured at the negative density. The four largest peaks correspond to the mercury positions, one each for the four subunits in the icosahedral asymmetric unit.

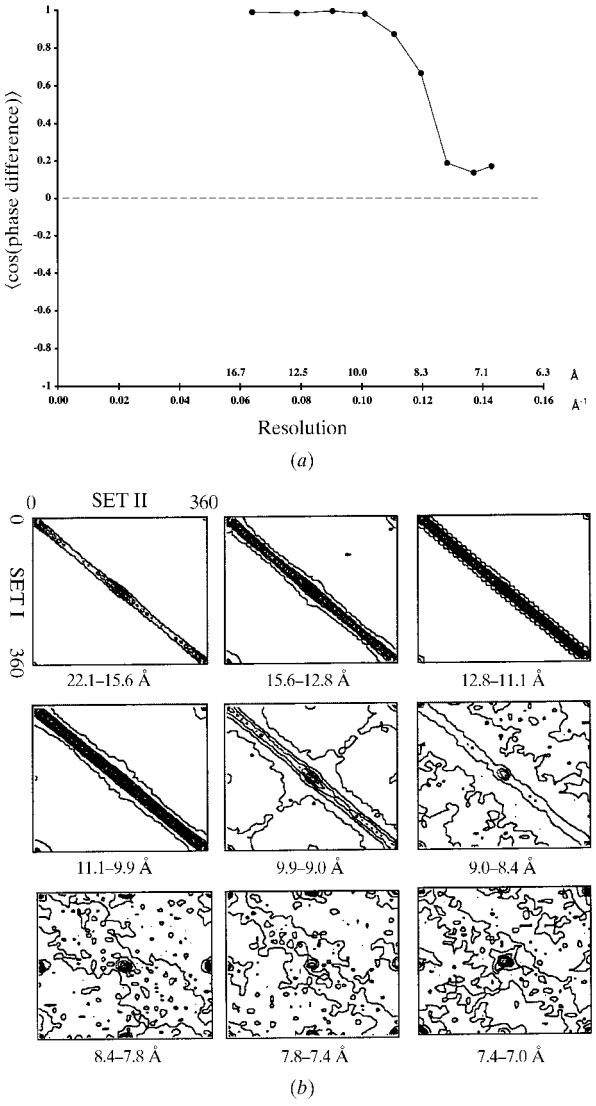


Fig. 7. (*a*) Plot of the mean cosine phase difference as a function of resolution. Set I and set II phases were compared with each other to 7 Å resolution. Limited correlation between the phases is observed beyond 8.5 Å resolution [⟨cos(phase difference)⟩ approaching 0.0]. (*b*) Comparison of the phases between set I and set II as a function of resolution. Plots show the frequency with which phases of set I correspond to phases of set II. Contours in the top two panels are drawn at an interval of 50, starting from 10. In the bottom panel, contours begin at 10 and are drawn at an interval of 15. The scattered distribution of the contours indicate limited correlation between the two phase sets beyond 10 Å resolution.

density for a number of residues, particularly amino acids with bulky side chains. A poly ala model was built into the electron-density map at 4 Å and the phases were computed to 3.5 Å resolution. The refined phases at 3.5 Å were used to compute a map that had well defined density for almost all of the ordered amino acids of the polypeptide. Much of the amino-acid sequence deduced from the gene sequence (Agrawal & Johnson, 1992) could be modeled into the map. The model was optimized by rigid-body refinement with *X-PLOR* (Brünger *et al.*, 1987). A new set of phases to 3.0 Å resolution was computed from the rigid-body refined model and further refined by four cycles of density averaging. All ordered parts of the map were modeled for each of the four subunits in the icosahedral asymmetric unit. Residues 61–595 and 58–590 were modeled for subunits *A* and *B*, respectively. Residues from 44–590 and 608–641 were modeled for subunit *C* and residues 44–590 and 616–641 were modeled for subunit *D*. Phases were next extended and refined to 2.8 Å resolution. The final overall molecular-replacement *R* factor was 18.4% and the correlation coefficient was 0.87 (Fig. 10). The final electron-density map was computed with 2465892 unique reflections between 20 and 2.8 Å resolution. A few cycles of conjugate-gradient-minimization refinement of the coordinates using *X-PLOR* and data between 8 and 3 Å (1048155 independent reflections, but only every second reflection was used in the refinement) resulted in a conventional crystallographic *R* factor of 27.4%. The polypeptide chain fold for the *C* subunit of NωV is shown in Fig. 11. The density for the first 43 amino acids of the polypeptide is disordered for the *C* and *D* subunits. For the *A* and *B* subunits the first ordered residues are 61 and 58, respectively. The N-terminus of the polypeptide is rich in basic amino acids (Agrawal & Johnson, 1992) and is presumably involved in interactions with the genomic RNA and may not maintain icosahedral symmetry. The C-terminal portion of the polypeptide chain is also rich in basic amino acids and is completely disordered for the *A* and *B* subunits. However, residues 608–641 are folded into an α-helix in the *C* and *D* subunits, with the basic side chains pointing towards the interior of the virus. Electron density for all four subunits showed clear breaks between the residues Asn570 and Phe571, the location of autocatalytic cleavage determined biochemically (Agrawal & Johnson, 1992, 1995). No ordered density could be attributed to the genomic RNA. A description of the structure and some of its functional implications was reported (Munshi *et al.*, 1996).

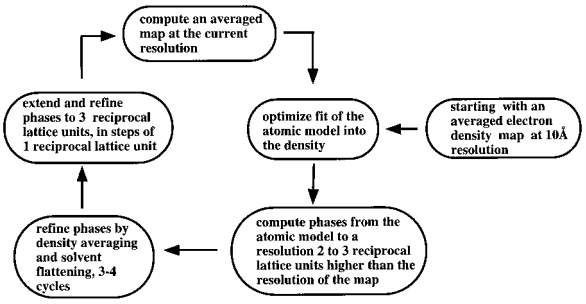


Fig. 8. Flow chart of the cyclic procedure used in computation and refinement of phases for the extended reflections.

11. Comparison of the phases obtained by conventional and unorthodox phase-extension procedure

A comparison of the final phase sets obtained by conventional phase-extension procedure, in which the phases for the extended reflections were computed by Fourier transformation of the averaged electron-density map, with the unorthodox procedure described in this paper showed that the phases between 10 and 5 Å resolution were significantly different from each other (Fig. 12). Many of the reflections between 10 and 8 Å corresponded to a related set of the four possible, with significant numbers of the reflections displaying phases that changed enantiomorph ($\alpha_C = -\alpha_U$) or were Babinet-inverted with changed enantiomorph ($\alpha_C = \pi - \alpha_U$). Phases between 8 and 7.4 Å were significantly correlated. Between 7.4 and 6 Å, most of the phases were of the alternate enantiomorph ($\alpha_C = -\alpha_U$), with a significant number of phases not correlated. Beyond 6 Å resolution, the phases were essentially not correlated. In the lower resolution range 20–17.5 Å a small number of reflections changed sign. The inconsistent phases at low

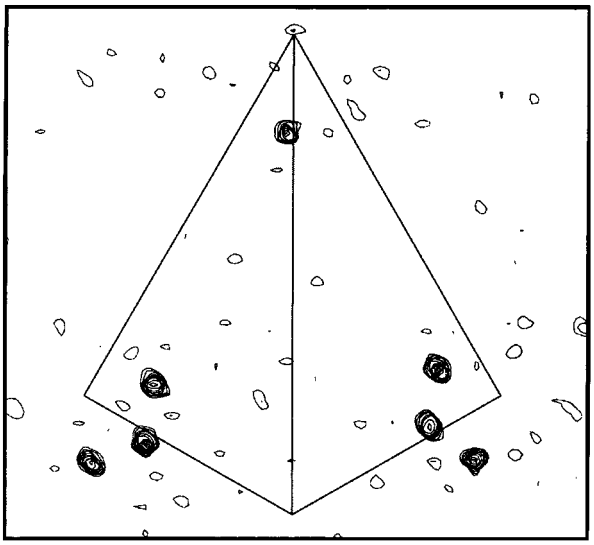


Fig. 9. The difference Fourier map between native and FMA derivative for data between 9 and 7 Å resolution represents a 2 Å slice at Z = 164 Å along the icosahedral threefold axis. The map was computed with phases refined at 10 Å resolution from the hollow spherical shell model M192 and extended to 7 Å using the unorthodox polypeptide-model-based approach. The four largest peaks correspond to the mercury sites, one for each of the four subunits in the icosahedral asymmetric unit (see Fig. 6).

resolution were probably due to the errors in the radii used for the spherical shell model. Comparison of the correlation coefficients to 5.2 Å resolution between the observed structure-factor amplitudes and the amplitudes calculated from the refined maps for the two phase-extension approaches is shown in Fig. 13. The overall mean correlation coefficients for data between 20 and 5.2 Å resolution for the conventional and unorthodox approaches were 0.854 and 0.867, respectively.

12. Discussion

The results presented indicate that the high-resolution crystal structure of a virus capsid with 60-fold non-crystallographic redundancy in the structure can be determined by starting from very approximate low-resolution phases computed from a hollow spherical shell of appropriate dimensions. However, under certain circumstances, the phase-extension process is prone to accumulate sets of phases of related, but unique, classes which are all compatible with the non-crystallographic symmetry constraints but give rise to different images (*e.g.* the two enantiomorphs of the object). The refinement process does not constrain the phases to a specific class in this case and different reflections form different phase subgroups. During the phase extension from 18 to 5 Å resolution in the *P*1 cell of NωV, reflections belonging to four different phase sets were observed. It could be argued that these four different phase sets already existed in the initial phases derived from the hollow spherical shell model. The final phase comparison in Fig. 12, however, suggests that only a very small number of reflections had phases with the opposite hand ($\alpha_C = -\alpha_U$). Also, the experiments carried out with a synthetic structure (model X) placed in the *P*1 cell of NωV clearly demonstrate that the phase crossovers to icosahedrally equivalent sets take place during extension.

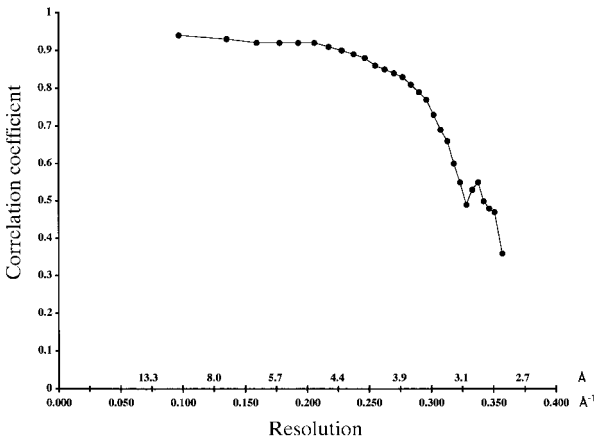


Fig. 10. Plot of correlation coefficient of the final averaged map as a function of resolution for data between 20 and 2.8 Å resolution.

The choice of phase set selected during phase extension has been shown to be sensitive to the molecular envelope used for density averaging (Valegård *et al.*, 1991). In practice, the molecular envelopes that can be determined from low-resolution electron microscopy images or from X-ray electron-density maps computed with phases from a homologous structure are not of sufficient detail to help constrain the handedness of phases globally. However, the shape of the molecular envelope has been shown to be less important in the structure determination of several virus structures (Speir *et al.*, 1995; Tsao *et al.*, 1992).

The structure determination of foot-and-mouth disease virus (Acharya *et al.*, 1989) posed many problems during phase extension, despite the starting phases being obtained from a relatively similar picornavirus capsid structure. Similarly the efforts to determine the high-resolution structure of satellite tobacco mosaic virus and satellite panicum mosaic virus starting from the known structure of satellite tobacco necrosis virus were unsuccessful (Ban & McPherson, 1995; Larson *et al.*, 1993). Difficulties were also observed in phase extension to high resolution, starting from SIR

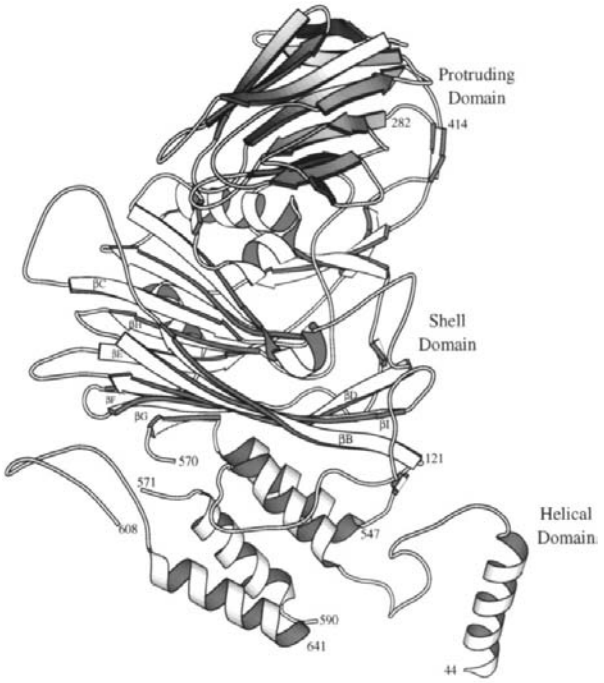


Fig. 11. Ribbon drawing (Kraulis, 1991) of the *C* subunit of NωV. The polypeptide is folded into three discrete domains. The central β-structure is an eight-stranded antiparallel β-barrel that forms the contiguous shell of the capsid. The two β-sheets of the barrel are labeled β-BIDG and β-CHEF consistent with the nomenclature used in describing the canonical β-barrel structure in other viruses. The helical domain facing the interior of the virus is made up of the N and C terminal segments of the polypeptide chain and contains the cleaved γ-peptide. The 12-stranded β-structure on the surface has the immunoglobin IgC type fold and is inserted between the strands βE and βF of the shell domain.

phases, in the structure determination of cowpea mosaic virus (Stauffacher *et al.*, 1987). In most of the virus structure determinations that have been difficult to solve by phase extension from low to high resolution, the virus particles occupied special positions in the crystal unit cell. Contrary to this, phase extension from very low (20 Å) resolution to high resolution has been relatively easy for those viruses that crystallized in space groups with virus particles packed in an acentric arrangement (Agbandje *et al.*, 1993; Luo *et al.*, 1987; McKenna *et al.*, 1992; Speir *et al.*, 1995).

It is not clear how the different phase sets are 'selected' during the extension process. The centric arrangement of the virus particles in the crystal probably contributed to the phase ambiguity during extension, although the phases themselves are not centric. In space groups such as *R*3, *R*32 and *P*1 with one virus particle in the unit cell, the relationship between the structure factors at the edge of the resolution is also limited owing to the size of the particle approaching the unit-cell size (Arnold & Rossmann, 1986). Thus, during extension of phases by Fourier transformation of the averaged map, the phases computed for the reflections in the next shell may assume near-random values which, upon refinement, converge to one of the four possible solutions.

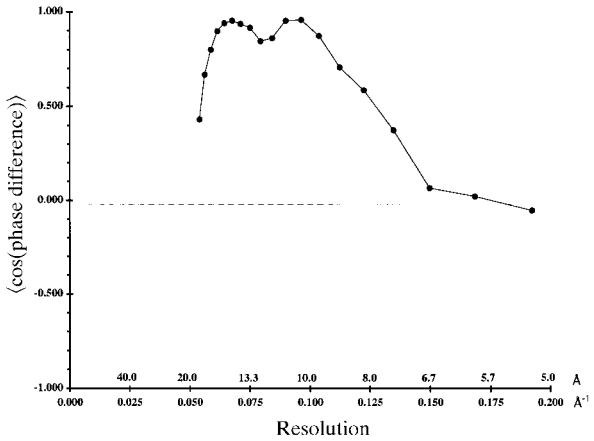


Fig. 12. Plot of the mean cosine phase difference for data between 20 and 5.2 Å resolution, between the final refined phase sets obtained by the two different approaches of phase extension. Poor correlation is observed beyond 10 Å resolution.

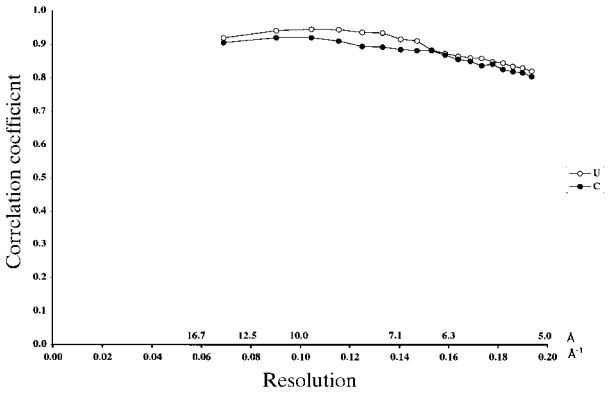


Fig. 13. Comparison of the correlation-coefficient profiles for data between 20 and 5.2 Å resolution, between the final refined phase sets obtained by the two different approaches of phase extension.

The strategy described in this paper can be successfully utilized in the structure determination of virus capsids, should the conventional phase-extension process fail to give an interpretable electron-density map. The procedure can be viewed as a distant relative of the '*Shake and bake*' approach to phase determination (Ealick, 1997). In the application described here the atom positions input for phase determination are not random, but they are not necessarily correct. They must be positioned with adequate accuracy to 'orient' the phases towards a consistent choice of the four possible phase sets conforming to the phase-refinement constraint of icosahedral symmetry. Thus, along with the limited requirements of the 'model' described above, our results show that the structure determination of a virus capsid with 60-fold non-crystallographic redundancy can be successfully initiated with *ab initio* phasing methods and extended to high resolution, thus avoiding the difficulties often encountered with the MIR method.

We wish to thank Michael Rossmann, Michael Chapman, Robert McKenna, the late Jin-bi Dai and Mathur R. N. Murthy for stimulating discussions; Michael Chapman for independently confirming the optimal inner and outer radii for the spherical-shell model; Bonnie McKinney, Tim Schmidt, Jean Cavarelli and Wu Bomu for help in crystallization and data processing; members of the Johnson's lab for data collection at the synchrotron sources; staff of CHESS and NSLS for their dedicated help; Jean-Pierre Wery for building the BBV-based $T = 4$ model; Marius Cornea-Hasegan, Bob Lynch, Dan Marinescu, Jodi Muckelbauer, Andrea Hadfield and Zhongyun Zhang for help with computation on the parallel processing machine, Andrzej Rajca for help with programming and the computing center staff of Purdue University for their dedicated help. This work was supported by a grant from the National Institutes of Health (R01 GM34220) to JEJ. This is manuscript number 11471-MB at The Scripps Research Institute.

References

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

Agbandje, M., McKenna, R., Rossmann, M. G., Strassheim, M. L. & Parrish, C. R. (1993). *Proteins Struct. Funct. Genet.* **16**, 155–171.

Agrawal, D. K. & Johnson, J. E. (1992). *Virology*, **190**, 806–814.

Agrawal, D. K. & Johnson, J. E. (1995). *Virology*, **207**, 89–97.

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

Ban, N. & McPherson, A. (1995). *Nature Struct. Biol.* **2**, 882–890.
Brünger, A. T., Kuriyan, J. & Karplus, M. (1987). *Science*, **235**, 458–460.
Cavarelli, J., Bomu, W., Liljas, L., Kim, S., Minor, W., Munshi, S., Muchmore, S., Schmidt, T. & Johnson, J. E. (1991). *Acta Cryst.* B**47**, 23–29.
Chapman, M. S., Tsao, J. & Rossmann, M. G. (1992). *Acta Cryst.* A**48**, 301–312.
Chen, Z., Stauffacher, C., Li, Y., Schmidt, T., Bomu, W., Kamer, G., Shanks, M., Lomonossoff, G. & Johnson, J. E. (1989). *Science*, **245**, 154–159.
Cornea-Hasegan, M. A., Zhang, Z., Lynch, R. E., Marinescu, D. C., Hadfield, A., Muckelbauer, J. K., Munshi, S., Tong, L. & Rossmann, M. G. (1995). *Acta Cryst.* D**51**, 749–759.
Ealick, S. E. (1997). *Structure*, **5**(4), 469–72.
Golmohammadi, R., Fridborg, K., Bundule, M., Valegård, K. & Liljas, L. (1996). *Structure*, **4**, 543–554.
Grant, R. A., Filman, D. J., Fujinami, R. S., Icenogle, J. P. & Hogle, J. M. (1992). *Proc. Natl Acad. Sci. USA*, **89**, 2061–2065.
Harrison, S. C., Olson, A. J., Schutt, C. E., Winkler, F. K. & Bricogne, G. (1978). *Nature (London)*, **276**, 368–373.
Hendry, D., Hodgson, V., Clark, R. & Newman, J. (1985). *J. Gen. Virol.* **66**, 627–632.
Johnson, J. E., Akimoto, T., Suck, D., Rayment, I. & Rossmann, M. G. (1976). *Virology*, **75**, 394–400.
Johnson, J. E., Munshi, S., Liljas, L., Agrawal, D., Olson, N. H., Reddy, V., Fisher, A., McKinney, B., Schmidt, T. & Baker, T. S. (1994). *Arch. Virol.* **9**, 497–512.
Jones, T. A. (1992). *CCP4 Study Weekend 1992: Molecular Replacement.* Warrington: Daresbury Laboratory.
Jones, T. A., Cowan, S., Zou, J. Y. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.
Kraulis, P. J. (1991). *J. Appl. Cryst.* **24**, 946–950.
Larson, S. B., Koszelak, S., Day, J., Greenwood, A., Dodds, J. A. & McPherson, A. (1993). *Nature (London)*, **361**, 179–182.
Liddington, R. C., Yan, Y., Moulai, J., Sahli, R., Benjamin, T. L. & Harrison, S. C. (1991). *Nature (London)*, **354**, 278–284.
Liljas, L., Unge, T., Jones, T. A., Fridborg, K., Lövgren, S., Skoglund, U. & Strandberg, B. (1982). *J. Mol. Biol.* **159**, 93–108.
Luo, M., He, C., Toth, K. S., Zhang, C. X. & Lipton, H. L. (1992). *Proc. Natl Acad. Sci. USA*, **89**, 2409–2413.
Luo, M., Vriend, G., Kamer, G., Minor, I., Arnold, E., Rossmann, M. G., Boege, U., Scraba, D. G., Duke, G. M. & Palmenberg, A. C. (1987). *Science*, **235**, 182–191.
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.
Munshi, S., Liljas, L., Cavarelli, W., Bomu, W., McKinney, B., Reddy, V. & Johnson, J. (1996). *J. Mol. Biol.* **261**, 1–10.
Olson, N. H., Baker, T. S., Johnson, J. E. & Hendry, D. A. (1990). *J. Struct. Biol.* **105**, 111–122.
Rayment, I. (1983). *Acta Cryst.* A**39**, 102–116.
Rayment, I., Baker, T. S., Caspar, D. L. D. & Murakami, W. T. (1982). *Nature (London)*, **295**, 110–115.
Rossmann, M. G. (1972). *The Molecular Replacement Method.* New York: Gordon & Breach.
Rossmann, M. G. (1976). *Acta Cryst.* A**32**, 774–777.
Rossmann, M. G. (1979). *J. Appl. Cryst.* **12**, 225–238.
Rossmann, M. G. (1990). *Acta Cryst.* A**46**, 73–82.
Rossmann, M. G., Arnold, E., Erickson, J. W., Frankenberger, E. A., Griffith, J. P., Hecht, H. J., Johnson, J. E., Kamer, G., Luo, M., Mosser, A. G., Rueckert, R. R., Sherry, B. & Vriend, G. (1985). *Nature (London)*, **317**, 145–153.
Rossmann, M. G. & Blow, D. M. (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.
Speir, J. A., Munshi, S., Wang, G., Baker, T. S. & Johnson, J. E. (1995). *Structure*, **3**, 63–78.
Stauffacher, C. V., Usha, R., Harrington, M., Schmidt, T., Hosur, M. V. & Johnson, J. E. (1987). *Crystallography in Molecular Biology*, edited by D. Moras, J. Drenth, G. Strandberg, D. Suck & K. Wilson, pp. 293–308. New York & London: Plenum.
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. (1991). *Acta Cryst.* B**47**, 949–960.
Yan, Y., Stehle, T., Liddington, R. C., Zhao, H. & Harrison, S. C. (1996). *Structure*, **4**, 157–164.