$$\Phi_{\pi'\sigma}(h00, \Psi)$$
$$= (o/s^2)\{C_{hx}[f_{12}\textstyle\sum_{12}(scS_\Theta C_{2\Psi} - s^2 C_\Theta S_\Psi)$$
$$+ f_{13}\textstyle\sum_{13}(sC_\Theta C_\Psi + cS_\Theta S_{2\Psi}) - f_{23}\textstyle\sum_{23} sS_\Theta C_{2\Psi}]$$
$$+ iS_{hx}[f_{12}\textstyle\sum_{2}(scS_\Theta C_{2\Psi} - s^2 C_\Theta S_\Psi)$$
$$+ f_{13}\textstyle\sum_{3}(sC_\Theta C_\Psi + cS_\Theta S_{2\Psi})$$
$$- f_{23}\textstyle\sum_{123} sS_\Theta C_{2\Psi}]\}. \qquad (A2)$$

0k0 reflections (k odd)

$$\Phi_{\sigma'\sigma}(0k0, \Psi)$$
$$= (o/s^2)\{C_{ky} f_{13}\textstyle\sum_{13}(sS_{2\Psi} - 2cC_\Psi^2)$$
$$+ iS_{ky} f_{13}\textstyle\sum_{123}(sS_{2\Psi} - 2cC_\Psi^2)\}, \qquad (A3)$$

$$\Phi_{\pi'\sigma}(0k0, \Psi)$$
$$= (o/s^2)\{C_{ky}[f_{12}\textstyle\sum_{12} C_\Theta(s^2 S_\Psi - scC_\Psi)$$
$$+ f_{13}\textstyle\sum_{13} S_\Theta(sC_{2\Psi} + cS_{2\Psi}) + f_{23}\textstyle\sum_{23} sC_\Theta C_\Psi]$$
$$+ iS_{ky}[f_{12}\textstyle\sum_{1} C_\Theta(s^2 S_\Psi - scC_\Psi)$$
$$+ f_{13}\textstyle\sum_{123} S_\Theta(sC_{2\Psi} + cS_{2\Psi})$$
$$+ f_{23}\textstyle\sum_{3} sC_\Theta C_\Psi]\}. \qquad (A4)$$

00l reflections (l odd)

$$\Phi_{\sigma'\sigma}(00l, \Psi)$$
$$= (o/s^2)\{C_{lz}[f_{12}\textstyle\sum_{12} sS_{2\Psi} - 2f_{13}\textstyle\sum_{13} cC_\Psi^2$$
$$- f_{23}\textstyle\sum_{23} scS_{2\Psi}]$$
$$+ iS_{lz}[f_{12}\textstyle\sum_{123} sS_{2\Psi} - 2f_{13}\textstyle\sum_{1} cC_\Psi^2$$
$$- f_{23}\textstyle\sum_{2} scS_{2\Psi}]\}, \qquad (A5)$$

$$\Phi_{\pi'\sigma}(00l, \Psi)$$
$$= (o/s^2)\{C_{lz}[f_{12}\textstyle\sum_{12} sS_\Theta C_{2\Psi}$$
$$+ f_{13}\textstyle\sum_{13}(-sC_\Theta C_\Psi + cS_\Theta S_{2\Psi})$$
$$- f_{23}\textstyle\sum_{23}(s^2 C_\Theta S_\Psi + scS_\Theta C_{2\Psi})]$$
$$+ iS_{lz}[f_{12}\textstyle\sum_{123} sS_\Theta C_{2\Psi}$$
$$+ f_{13}\textstyle\sum_{1}(-sC_\Theta C_\Psi + cS_\Theta S_{2\Psi})$$
$$- f_{23}\textstyle\sum_{2}(s^2 C_\Theta S_\Psi + scS_\Theta C_{2\Psi})]\}. \qquad (A6)$$

References

BUSING, W. R. & LEVY, H. A. (1967). *Acta Cryst.* **22**, 457-464.
DMITRIENKO, V. E. (1983). *Acta Cryst.* A**39**, 29-35.
DMITRIENKO, V. E. (1984). *Acta Cryst.* A**40**, 89-95.
FANCHON, E. & HENDRICKSON, W. A. (1990). *Acta Cryst.* A**46**, 809-820.
International Tables for Crystallography (1983). Vol. A. Dordrecht: Kluwer Academic Publishers.
KIRFEL, A. & PETCOV, A. (1991). *Z. Kristallogr.* **195**, 1-15.
KIRFEL, A. & PETCOV, A. (1992). *Acta Cryst.* A**48**, 247-259.
KIRFEL, A., PETCOV, A. & EICHHORN, K. (1991). *Acta Cryst.* A**47**, 180-195.
TEMPLETON, D. H. & TEMPLETON, L. K. (1980). *Acta Cryst.* A**36**, 237-241.
TEMPLETON, D. H. & TEMPLETON, L. K. (1982). *Acta Cryst.* A**38**, 62-67.
TEMPLETON, D. H. & TEMPLETON, L. K. (1985). *Acta Cryst.* A**41**, 133-142, 365-371.
TEMPLETON, D. H. & TEMPLETON, L. K. (1986). *Acta Cryst.* A**42**, 478-481.
TEMPLETON, L. K. & TEMPLETON, D. H. (1988). *Acta Cryst.* A**44**, 1045-1051.
WILLIS, B. T. M. & PRYOR, A. W. (1975). *Thermal vibrations in Crystallography*. Cambridge Univ. Press.

Acta Cryst. (1993). A**49**, 45-55

Methods Used in the Structure Determination of Foot-and-Mouth Disease Virus

BY ELIZABETH FRY, RAVINDRA ACHARYA* AND DAVID STUART

Laboratory of Molecular Biophysics, Rex Richards Building, South Parks Road, Oxford OX1 3QU, England

(*Received* 14 *March* 1992; *accepted* 13 *July* 1992)

Abstract

The structure of foot-and-mouth disease virus (FMDV) strain O_1 BFS 1860 has been determined to 2.9 Å resolution using the molecular-replacement method [Acharya, Fry, Stuart, Fox, Rowlands & Brown (1989). *Nature (London)*, **337**, 709-716]. Crystals of the virus with average dimensions 0.12 × 0.06 × 0.12 mm belong to space group *I*23, $a = 345$ Å with 1/12 of the icosahedral particle per asymmetric unit giving fivefold noncrystallographic redundancy. Oscillation diffraction photographs were collected at the SERC Synchrotron Radiation Source at Daresbury in accordance with strict disease security regulations. The ambiguity in particle orientation was resolved using a self-rotation function and starting estimates of the phases to 8 Å were derived from the known structures of two picornaviruses similarly oriented in the *I*23 unit cell. The phases were refined

* Present address: Department of Biochemistry, 4-West, University of Bath, Claverton Down, Bath BA2 7AY, England.

0108-7673/93/010045-11$06.00

and extended using iterative averaging and solvent flattening with the implementation of a simple automatic envelope-determination procedure to increase the phasing power available.

Introduction

Foot-and-mouth disease viruses constitute the *Aphthovirus* genus of the Picornaviridae. These animal viruses are spherical (approximately 300 Å in diameter) and have a molecular weight of the order of 8.5×10^6 dalton. They comprise a 30–50 Å thick protein shell (capsid), assembled from 60 copies of each of four proteins VP1–VP4 arranged with icosahedral symmetry. This viral coat surrounds and protects the genome, a single strand of RNA.

Whilst structural redundancy has not yet yielded a truly *ab initio* phase determination, much of its potential has been exploited in virus crystallography. Prior to this crystallographic analysis, the crystal structures of three other picornaviruses were known, a member from each of the *Rhinovirus* [human rhinovirus 14 (HRV14) (Rossmann *et al.*, 1985)], *Enterovirus* [poliovirus Mahoney strain (Hogle, Chow & Filman, 1985)] and *Cardiovirus* [mengovirus (Luo *et al.*, 1987)] genera. The structures of both HRV14 and poliovirus were determined using isomorphous-replacement starting phases (based on poor low-resolution heavy-atom derivatives) and noncrystallographic symmetry averaging to improve and extend these to high resolution. This success was taken a step further in the solution of mengovirus (Luo *et al.*, 1987) by molecular replacement based on the crudely similar HRV14 structure. In this case, 60-fold redundancy provided a powerful phase constraint. With only fivefold redundancy for FMDV, a strategy similar to that used for mengovirus failed and a molecular replacement solution was only possible by devising a very careful envelope-determination strategy to maximize the phasing power from the solvent contribution. The strategy described in this paper has since proved successful in an analogous situation for the structure determination of carnation mottle virus (Morgunova *et al.*, 1992). Other recent crystallographic analyses have been performed on unrelated viruses (Valegård, Liljas, Fridborg & Unge, 1990; Tsao *et al.*, 1991; Liddington *et al.*, 1991; McKenna *et al.*, 1992). These studies have all used some information from heavy-atom-derivative data, thus the structure determination described here remains unusual in succeeding from poor molecular-replacement starting phases and minimal non-crystallographic symmetry averaging.

In brief, the orientation of the virus particle was determined using a self-rotation function. Phases to 8 Å were then obtained from a hybrid of the known structures of HRV14 and mengovirus. The phases were improved using the constraints of noncrystallographic symmetry and solvent flattening *via* a special adaptation of the Bricogne real-space averaging procedure (Bricogne, 1976). We will describe in detail only those aspects where our procedure differs from those previously documented.

Data collection

Cubic crystals of cloned virus were obtained by dialysis aginst 17% $(NH_4)_2SO_4$ in 0.1 *M* PO_4, pH 7.6 at room temperature (Fox *et al.*, 1987). Isomorphous crystals of a variant virus were also grown. These achieved the same average dimensions ($0.12 \times 0.06 \times 0.12$ mm) in a few days compared with the two to four weeks for the parent virus. The *I*23 unit cell has a cell edge $a = 345$ Å. With a virus particle positioned at (0, 0, 0) (on a point of 23 symmetry) there are two particles in the unit cell. The asymmetric unit is 1/12 of the virus particle and there is fivefold noncrystallographic redundancy. The packing density in the crystal V_m is 2.7 Å^3 dalton^{-1}. The approximate volume of solvent plus RNA (disordered) is 72% of the unit cell.

The large unit cell, small size and extreme radiation sensitivity of these crystals dictated the use of synchrotron radiation for its beam geometry and intensity. For each trip made to the SERC synchrotron at Daresbury (SRS), permission was obtained from the Ministry of Agriculture, Fisheries and Food to transport live virus crystals from the containment laboratory of Coopers Animal Health (formerly Wellcome Biotech) at Pirbright, Surrey. This was conditional on the continual presence of a disease security officer and strict adherence to an agreed protocol.

Data were recorded on small-angle (0.4°) rotation photographs (one per crystal) using the Enraf–Nonius oscillation camera. The crystals diffracted to at least 2.9 Å for a typical exposure time of 15 min. Most data were recorded at the fixed temperature of 288 K. To test the effect of lower temperature, a crystal was cooled to 277 K; the diffraction pattern clearly indicated disordering of the crystal lattice, which was, however, reversed by warming the crystal to 288 K. This project was one of the first to employ exclusively short wavelengths <1 Å. Comparative tests involving the collection of data using $\lambda = 1.488$ Å (station 7.2) and approximately $\lambda = 0.90$ Å (station 9.6) showed considerable improvement not only in the signal-to-noise ratio for comparable crystals but also in the crystal lifetime.

Once mounted in 0.7 mm quartz capillaries (disease security regulations required the use of quartz), most crystals preferentially aligned such that a $[\bar{1}0\bar{1}]$ zone axis was perpendicular to the capillary wall. With this axis along the X-ray beam, there was no guarantee that any given axis was aligned with the rotation axis. Optical alignment was used to prealign the crystals to some extent but their fragility prevented reorienta-

tion in the quartz tube. The neglect of setting operations is common with virus crystals prone to radiation damage (Rossmann *et al.*, 1985; Fry, Logan & Stuart, 1992).

The films were digitized on a 50 μm raster on the Joyce Loebl Scandig III rotating-drum microdensitometer over an optical density range of 0–2, the data being recorded on nine-track magnetic tapes. Whilst most of the films used in the structure determination were indexed manually, an autoindexing procedure based on the *XDS* algorithm of Kabsch (1988), modified to enable the manipulation of the film image in a routine fashion, is now used reliably. The images were analysed using a specially adapted version of the integration program *MOSCO*9 (Stuart, Levine, Muirhead & Stammers, 1979; Stuart, unpublished). With various zones present on each film, there is considerable variation in the size, shape and spacing of the reflections, which makes it impossible to specify a single set of spot and background parameters that are optimal for each spot across the whole film. In addition, the limited dynamic range of the films (in many cases the background was almost saturated) coupled with the substantial number of overlapped reflections, especially at higher resolution, resulted in the majority of reflections being discarded on the basis of the standard rejection criteria. Thus, with the film image stored in memory, bit maps were created to indicate whether a pixel was used in one or more spots. To enable measurement of overlapped spots, only pixels not within another spot were profile-fitted. Any background pixels overlapping a spot were automatically excluded. Thus, for accurate background estimation in the vicinity of a spot, large background areas could be defined with the program tailoring the exact area used. Furthermore, with the omission of saturated pixels from spot profiles, it was possible to profile-fit saturated spots. To reduce the likelihood of misindexing for film packs lacking a direct-beam mark, a convolution search was implemented in the program to determine the correct camera constants. The previously implemented learning procedures for positioning spot windows and defining spot-intensity measurement profiles worked well (Wilson *et al.*, 1983).

Data used in the structure determination were initially processed to 4.5 Å resolution to enable refinement of the crystal-orientation and rocking-curve parameters before going to higher resolution and to provide a data set to resolve the ambiguity of the orientation of the particle in the cubic unit cell. This ambiguity arises because icosahedral particles do not possess fourfold symmetry. Thus, there are two different ways of placing an icosahedron on a specific threefold axis, which are related by a 90° rotation about a particle twofold axis. If Friedel's law holds, we may bring all data to a common orientation by exchanging the h and k indices as required. A subset of film packs was accumulated showing good agreement without any indices swapped. Subsequent packs were tested for agreement against this reference set using a convolution-search method over a misindexing grid, calculating correlation coefficients between the intensities of reflections from the reference set of data and those from the new pack, with both possible choices of indices.

Table 1. *Summary of data*

The total number of observed reflections to 2.9 Å was 393 699 and the mean redundancy was 3.0. R_{merge} on all data (intensities) was 13.9%.

Resolution range (Å)	Unique observed reflections ($F > 3\sigma$)	Percentage of possible observations
10 000–25	182 (175)	67.9
25–15	856 (812)	95.4
15–10	2594 (2465)	97.1
10–7.5	4909 (4569)	96.5
7.5–5.0	20020 (17864)	96.7
5.0–3.5	53634 (46716)	95.9
3.5–3.0	42644 (30754)	85.7
3.0–2.9	5661 (2806)	39.3
2.9–2.8	1703 (662)	10.3
Total (to 2.9)	130500 (106161)	87.2

Post-refinement (Schutt & Winkler, 1977) greatly improved the quality of the data, thus providing an accurate classification of partially and fully recorded reflections, which is essential with data collected for a narrow φ range with only one exposure per crystal. The rocking curve for reflections passing through the Ewald sphere was modelled by an isotropic cosine profile. It was usual to fix the cell dimensions and angles and refine the wavelength, misorientation angles and mosaicity. All the fully recorded reflections measured were included in the final data set (Table 1) since even the weakest reflections provide indirectly important restraints on the phases of the stronger data ($R_{merge} = 13.9\%$), where

$$R_{merge} = \left(\sum_h \sum_i | I_h - I_{hi} | / \sum_h \sum_i I_{hi} \right) \times 100, \qquad (1)$$

I_h is the weighted-mean measured intensity of the observations I_{hi}. Negative intensity measurements were treated using Bayesian statistics (French & Wilson, 1978).

In all, over 500 crystals were mounted and taken to the SRS, from which 147 usable data photographs were obtained. Data from both the parent (60 film packs) and variant virus (46 film packs) were combined to constitute a virtually complete set of data for the structure determination. If the effect of random errors in the data is removed, then the residual mean fractional isomorphous difference between the variant and native data does not exceed 5%. These data were subsequently separated and used for the calculation of difference maps (Parry *et al.*, 1990).

Particle orientation and position

The ambiguity in particle orientation was resolved by the calculation of a self-rotation function (POLARRFN) using the Kabsch (1988) implementation of the adaptation (Tanaka, 1977) of Crowther's (1972) rotation-function algorithm to work directly in spherical polar coordinates. All the peaks expected from the icosahedral symmetry were clearly visible on the $\kappa = 72°$ (fivefold), $\kappa = 120°$ (threefold) and $\kappa = 180°$ (twofold) sections. Comparison of the $\kappa = 72°$ section (Fig. 1*a*) with Figs. 1(*b*) and (*c*) (showing the two possible orientations relative to the unit-cell axes), demonstrates that the orientation is as shown in Fig. 1(*c*). This is the same virion orientation as in HRV14 coordinates deposited in the Protein Data Bank (Bernstein *et al.*, 1977). Since the position of the particle centre is defined by the intersection of crystallographic twofold and threefold axes, we may now directly calculate starting phases.

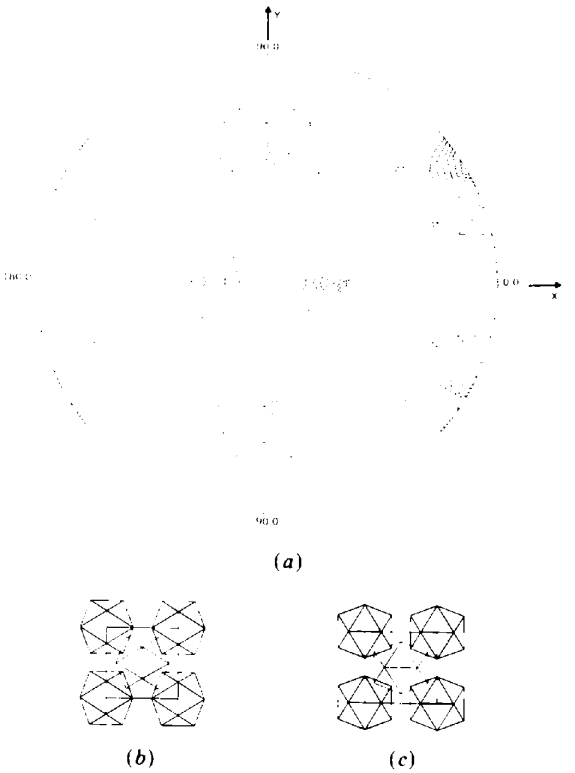


Fig. 1. Interpretation of the self-rotation function. (*a*) A plot of the self-rotation function for FMDV section $\kappa = 72°$. The data used for this calculation were from film packs processed to 4.5 Å resolution constituting 96% of the unique reflections. Resolution cutoffs of 9 and 6 Å were chosen to reduce virus/RNA and virus/solvent effects and thus exhibit clearly the capsid symmetry. The data were sharpened by applying a temperature factor of -20 Å^2. Only a small number of very weak reflections were omitted from the calculation. The integration radius was limited to 30 Å by the number of Bessel functions in the program; ideally, this would have been considerably larger. The noncrystallographic fivefold peaks are 65% of the origin peak height and no noise peaks are higher than 14%. (*b*), (*c*) The two possible orientations for icosahedra relative to the unit-cell axes.

Structure-factor amplitudes were calculated for both HRV14 and mengovirus placed in the FMDV cell (the mengovirus coordinates were a gift from Professor M. G. Rossmann prior to deposition in the Protein Data Bank) and those for the composite virus 'RHENGO' were obtained as a simple average of these. An equal-weight strategy was adopted because, although sequence comparisons suggested a closer similarity with mengovirus than with HRV14, the coordinates for HRV14 had been refined against crystallographic data. However, both viruses are divergent from FMDV; the percentage of residues of VP1 that match FMDV in position and sequence is 14.3% for mengovirus and 11.4% for HRV14. An R factor,

$$R = \left(\sum_h \| F_{h,\mathrm{obs}} | - | F_{h,\mathrm{calc}} \| / \sum_h | F_{h,\mathrm{obs}} | \right) \times 100, \quad (2)$$

was then calculated between the observed and calculated structure-factor amplitudes. Initial results showed very poor low-resolution R factors (see Fig. 4), which were mostly attributable to poor solvent modelling. With masks used to define the solvent and RNA boundaries, the appropriate electron-density value was substituted into these regions. Individually generated RNA and solvent structure factors (appropriately scaled and B-factor attenuated) were then summed vectorially with the protein structure factors to achieve a simple bulk-solvent correction procedure. The agreement of structure-factor amplitudes from HRV14, mengovirus, 'RHENGO' and a model based on the predicted FMDV structure (Luo, Rossmann & Palmenberg, 1988) was assessed. The composite virus structure gave the best overall agreement. The agreement was still satisfactory at 8 Å, which was chosen initially to be the high-resolution cutoff for phase calculation. The 'RHENGO' structure factors scaled to the observed FMDV data with an R factor of 50% for all the data in the range ∞–8 Å. Deliberate misorientation of the virion by 90° increased R to 54%.

Phase improvement and extension by molecular-repacement averaging

Phase improvement was achieved largely using established programs (Bricogne, 1976). The averaging portion in this procedure is similar in many respects to that used in the determination of the structure of tobacco bushy stunt virus (TBSV) (Olson, Bricogne & Harrison, 1983) and we will address in detail only areas of difference. In the description that follows we refer to several portions of the crystallographic unit cell using terms illustrated in Fig. 2 (see also Olson, Bricogne & Harrison, 1983). Briefly, U_{532} corresponds to an icosahedral subunit, which is 1/60 of the virion.

U_{I23} defines an asymmetric unit in space group $I23$ and U_{I222} similarly in space group $I222$. When used in the context of real space, the term 'index' refers to the coordinates defining the position of a given pixel of electron density.

The electron density was averaged over the non-crystallographically related subunits within the viral envelope, the average electron density was back-transformed and then the resultant calculated phases were reapplied with suitable weighting to the original F_{obs} for the computation of a new electron-density map. On each cycle, the electron-density values for pixels representing the disordered interior portion of the virus and the disordered external solvent region were separately flattened to their mean values. This cyclical procedure was continued until convergence on a dedicated MicroVAX II computer and 440 Mbyte disk.

Fourier syntheses were performed in space group $I222$ since this is more easily implemented in the FFT algorithm. Two Fourier syntheses were performed in every cycle. On cycle 0, the electron-density maps were calculated with amplitude coefficients $2|F_{obs}| - |F_{RHENGO}|$, on later cycles (including the phase-extension process), the coefficients were simply $|F_{obs}|$. A finely sampled Fourier electron-density map was used as a source for averaging and calculated at approximately one-fifth of the resolution of the data to reduce errors introduced by the linear interpolation (Bricogne, 1976). In this case, linear interpolation was used and averaging was performed over the whole U_{I23} asymmetric unit. Triplication of this U_{I23} domain, followed by folding back into the $I222$ asymmetric unit, gives the $\bar{U}_{I222}$ domain. Thus, the interpolated densities were expanded to the $\bar{U}_{I222}$ domain prior to density modification. At this stage, an adapted version of *MODIFY* (Bricogne, 1976) (which accumulates and averages symmetry-related densities) was developed. These modifications allowed the use of an envelope (defining the U_{532} domain) that actually slightly overlapped adjacent icosahedral subunits, without introducing errors. Densities not averaged (solvent or RNA) were subsequently flattened.

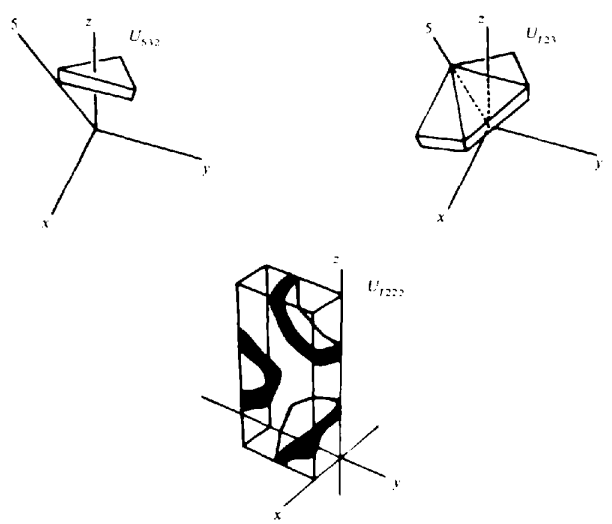


Fig. 2. The unique portions U_{532}, U_{I23} and $\bar{U}_{I222}$ used in various stages of the protocol are illustrated.

Phases for computing the next electron-density map were obtained from the calculated structure factors from the previous cycle, reduced to the U_{I23} asymmetric unit and combined with matching observed amplitudes. Whilst it is generally a good idea to include calculated structure factors to fill gaps in the sampling of the transform, these were not included during the first cycles of phase refinement to avoid bias from the starting model. In addition, reflections were weighted according to the error in their phases based on the fit of the calculated amplitudes of the phasing model to the observed amplitudes (Rayment, 1983). Near the resolution limit, calculated values were generally smaller than their correct values and were thus downweighted. After a few cycles of Rayment weighting, a unitary weighting scheme was used to allow the phases to settle (Rayment, 1983). Calculated and observed structure factors were scaled in resolution shells with the inclusion of various rejection criteria to prevent bias from spurious measurements.

(i) For reflections at lower resolution than 30 Å, F_{obs} was set to F_{calc}, where F_{obs} is the observed structure amplitude of reflection h and F_{calc} is the calculated structure amplitude of reflection h as a result of back-transforming the averaged and modified electron-density map.

(ii) Reflections where $|F_{obs} - F_{calc}| > A\sigma F_{obs}$, where A is an imposed σ cutoff value (usually 6), were rejected (treated as unobserved).

(iii) Reflections for which $|F_{obs} - F_{calc}| > D\bar{F}_{obs}$ were rejected, where $\bar{F}_{obs}$ is the mean value for that resolution shell and D is a specified cutoff value, usually about 0.8.

The total number of reflections rejected was monitored and was less than 0.5% of the total.

At this stage, the quality and progress of the phase determination were assessed by the calculation of correlation coefficients, C and R factors [from (2)] between the observed data and the calculated amplitudes from the constrained map as a function of $(\sin\theta)/\lambda$ and $|F_{obs}|$,

$$C = \sum_h (\langle F_{obs}\rangle - |F_{h,obs}|)(\langle F_{calc}\rangle - |F_{h,calc}|) \times \left[\sum_h (\langle F_{obs}\rangle - |F_{h,obs}|)^2 \sum_h (\langle F_{calc}\rangle - |F_{h,calc}|)^2\right]^{-1/2}. \quad (3)$$

In addition, the phase changes from cycle to cycle were tabulated to check for convergence.

Table 2. *Phase refinement at* 8 Å *resolution*

Expressions for R and C are given in (2) and (3), respectively. CPU time per cycle using a MicroVAX II RC is approximately 33 min.

Cycle number	Final R	Final C	Weighting (mean weight)
1	0.364	0.694	Rayment (0.687)
2-4	0.27	0.788	Rayment (0.746)
5-7	0.241	0.825	Unit
8-10	0.219	0.852	Unit†
11-19‡	0.212	0.857	Unit†

† Addition of unobserved data to help reduce series-termination errors.
‡ Convergence achieved.

A cycle of averaging with indices generated for the U_{123} domain in the absence of an envelope was performed prior to the envelope determination. The starting envelope was similar to that used in a number of other virus structure determinations (Rossmann *et al.*, 1985; Luo *et al.*, 1987) defined as the void between an inner sphere (radius 103.5 Å) and an outer sphere (radius 157 Å) truncated by planes normal to and at the midpoint of the vectors joining the centres of adjacent viruses. The inner boundary was possible since the viral RNA does not exhibit icosahedral symmetry. A pentameric portion was defined by specifying a radius (94.8 Å) relative to the correct fivefold axis, creating a 'truncated conical cap'. These bounds were chosen by inspection of an averaged but not truncated map.

The initial phases were improved by 19 cycles of averaging (Table 2) at 8 Å resolution. The R factor dropped from 50% before averaging to 21.2% after averaging, with a correlation coefficient of 0.857. The averaged maps showed good contrast between the regions expected to contain ordered protein (within a radius 100–150 Å) as opposed to the solvent and RNA. There was also some evidence for β-barrel structures within the protein region.

Phase extension

With confidence in the phases of the low-resolution data, a phase-extension strategy was embarked upon whereby the phases from a thin shell of structure factors just outside the resolution sphere of the previous cycle of refinement were taken and included with their corresponding observed amplitudes in the next cycle. Assuming a simple spherical model of the virus, the first positive region of the G function (Rossmann & Blow, 1962) extends to 1.6 reciprocal-lattice units from the origin, suggesting that the thickness of the extension shell should not exceed this. However, contrary to previously adopted procedures in which the additional phases were refined for a number of cycles before adding further terms (Rossmann *et al.*, 1985; Luo *et al.*, 1987), a fine-grained continuous extension procedure was used, extending the resolution in very small steps (one third of a reciprocal-lattice unit) on each cycle.

Calculated structure-factor amplitudes for the new data were generally weak and thus assigned to a separate 'resolution' shell for scaling. Rayment weighting was applied to the two outermost shells. The newly added data (except the outermost shells) included calculated values for the unmeasured structure factors to aid convergence. These were down-weighted by a factor of 0.5 to avoid bias.

After phase extension to 5.5 Å, the averaging R factor remained at 20%. Extra cycles of averaging at fixed resolution applying first Rayment weights to all the data and then unit weights produced no improvement. Furthermore, an averaged map computed with these phases did not make chemical sense. At this stage it became painfully obvious that the phase-extension procedure had failed. Since the noncrystallographic constraints were far less than in the only other comparable successful application of this procedure (Luo *et al.*, 1987), we decided to look for another means of increasing the phasing power. In particular, whilst a simple spherical envelope had been satisfactory for phase refinement at low resolution, it seemed likely that a more detailed envelope might be required to provide extra constraints during phase extension. An assessment of the current envelope showed that the total number of pixels outside the protein envelope (RNA and solvent) was only 37% of the unit-cell volume. However, the true total volume occupied by these two components was calculated to be 72% of the unit cell (based on a protein volume of 1.2 $Å^3$ $dalton^{-1}$). This indicated almost a 50% loss in the potential phasing power (assuming this to be directly related to the solvent content).

The strategy used for tightening the envelope

The procedure was based in part on the method developed by Wang (1985) and is shown schematically in Fig. 3. A map was calculated with data to a resolution of 8 Å, this was the maximum resolution to which reasonably reliable phases were available. In subsequent envelope determinations, higher-angle terms were included. This map was truncated, setting negative electron-density values in the protein region to zero. Structure factors calculated by Fourier back-transforming this map were multiplied by a B factor of 5000 $Å^2$ and used to calculate a smeared map at 8 Å resolution. The formerly defined envelope was used to define 'cautious' outer and inner envelopes between which all the protein density would be expected to occur. Within the potentially proteinacious region thus created, the electron-density values were accumulated into a histogram (Fig. 3). The number of extra 'solvent' pixels required to tighten the originally defined envelope was gauged from the esti-

mated actual number of solvent pixels minus the number of solvent pixels defined by the cautious envelope. The appropriate cutoff point was then determined from the histogram. In practice, the number of extra pixels assigned to solvent was kept substantially below the number that could theoretically be added in and varied according to the phasing power required. R factors and correlation coefficients were checked with the tighter envelope for evidence of truncation (Rayment, 1983). Recalculation of the envelope was performed at intervals dependent on the convergence of the phases (in all, forty were calculated). Comparisons between sequentially generated envelopes showed the number of pixels changed to be always less than 3% of the total.

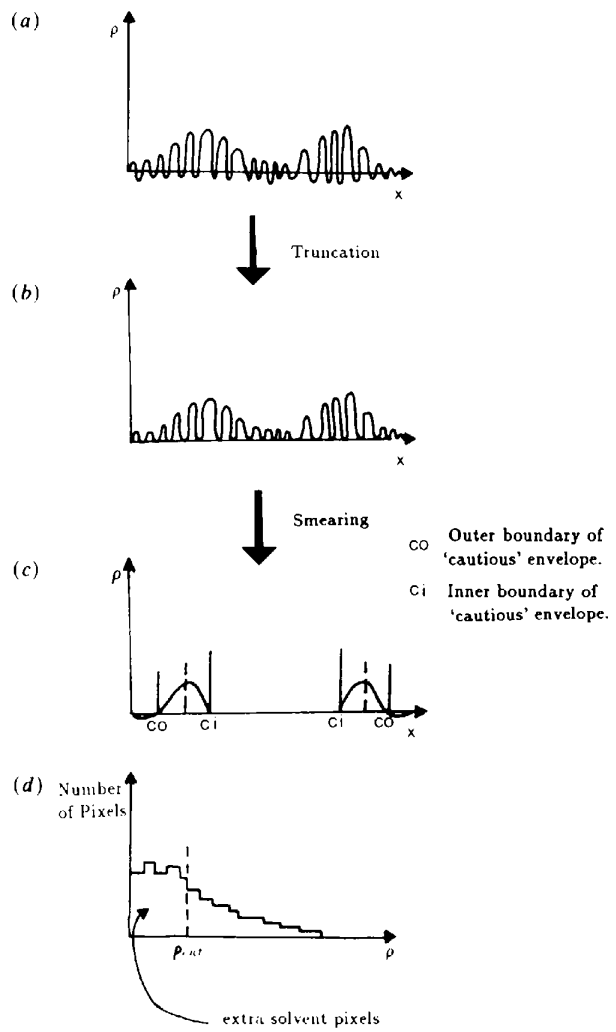


Fig. 3. Automatic envelope determination. The steps involved in envelope determination are depicted schematically. The starting map, (*a*), is truncated to (*b*). Structure factors from the back-transformed truncated map are multiplied by a B factor and used to calculate a smeared map (*c*). Electron-density values within the boundaries of the cautious envelope are accumulated into a histogram (*d*). The appropriate electron-density cutoff point is determined from the estimated actual number of solvent pixels minus the number of solvent pixels defined by the cautious envelope.

At this stage, other measures were introduced to deal with the low-angle data. The incompleteness of this shell of data and hence its domination by F_{calc}'s resulted in the phases failing to converge to their correct values. To avoid low-resolution ripples resulting from errors in these data, a low-angle cutoff was imposed initially. Then an improved procedure was introduced to reduce series-termination errors during the envelope calculation and to eliminate the possibility of any partially ordered structure within the capsid (arising from the RNA) distorting the envelope calculation: the data were modified by a Gaussian. This was executed prior to the calculation of the map to be truncated by attenuating the structure factors:

$$F'_{h,k,l} = \text{Atten} \times F_{h,k,l},$$

where

$$\text{Atten} = 1 - \exp[-B1(\sin^2\theta)/\lambda^2].$$

Then, after inversion of the truncated map, the structure factors were again attenuated, this time by the function

$$\text{Atten} = \exp[-B2(\sin^2\theta)/\lambda^2].$$

$B1$ and $B2$ were chosen by trial to be 400 and 1000, respectively (these values worked well but may not be optimal in other situations).

As a further means of improving the phases controlling the low-resolution structure, a bipartite 'Wang' envelope was constructed to allow the solvent and RNA regions to be averaged separately. The solvent and RNA regions were then assigned to their respective mean densities. In addition, the program truncated negatives within the protein region according to an electron-density cutoff ρ_{cutoff}, defined as

$$\rho_{cutoff} = (S\rho_{max} - \bar{\rho}_{solv})/(1 - S),$$

where ρ_{max} is the maximum protein electron density ($\simeq$6000), $\bar{\rho}_{solv}$ is the mean solvent electron density ($\simeq -60$) and S is a θ-dependent constant adjusted to give optimum convergence (Wang, 1985):

$$S = kd^2.$$

In practice, k was adjusted to keep the percentage of points truncated small (approximately 1% of the total map); increasing this percentage made the process unstable. For much of the work, the value chosen for k was $\sim$0.01.

As expected, the mean electron density in the external solvent converged to a value somewhat lower than that in the more electron-rich virus interior.

Successful phase extension

The resolution limit was reduced back to 8 Å and the phase-extension process started afresh, incorporating the new envelope-determination procedure and treatment of the low-resolution terms but otherwise using the same protocol as before. Since the RHENGO structure factors agreed reasonably with the FMDV measured data up to 5.5 Å resolution and the averaging R factor was 36% after only one cycle of averaging, some phase information from the RHENGO structure factors was used to 6 Å resolution. Thus, for newly phased reflections, the phase angle was taken as the mean of those derived from phase extension and from the RHENGO model. Throughout, a policy of cycling first with Rayment weights (applied to all the data) then with unit weights followed by envelope redetermination was adhered to.

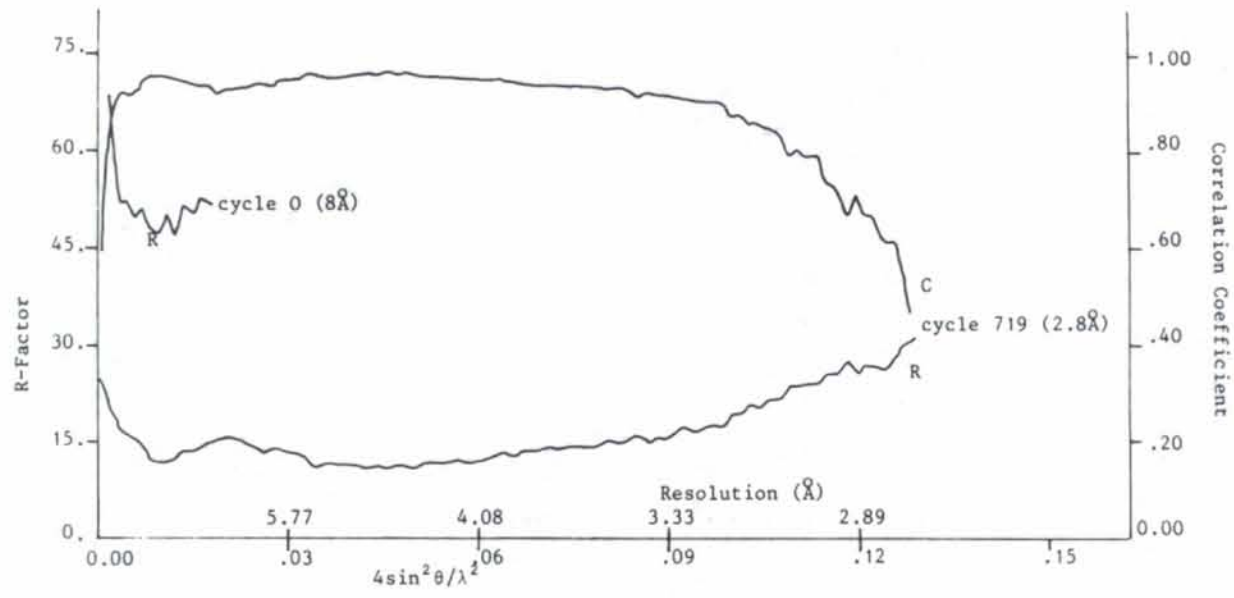
At 6 Å resolution a skew-averaged map corresponding to the icosahedral subunit was calculated. This made significantly more chemical sense than the equivalent map calculated during the aborted phase-extension attempt. A comparison of the phases with those from the initial attempt at phase extension showed the average difference to be 75°.

The map was checked against that of RHENGO for bias. The latter had considerably more density around the fivefold axis. A map generated at 4.9 Å resolution after further phase extension enabled the layers in the β-sheet to be resolved. Helix turns and loops could also be traced. Superimposing the HRV14 and mengo coordinates showed substantial shifts in the β-sheet. Overall, the electron density was closer to the mengovirus structure, though considerably different in places. The structure was evidently more compact than mengovirus or HRV14, thus providing reassurance that there was no residual bias from the starting phases. At this stage it was clear that the new procedures had succeeded and that the extended phase set was essentially correct. To continue the phase extension beyond 4.5 Å, the sampling grid was set to 0.67 Å, which was judged to be sufficiently fine to be satisfactory at the limit of the observed data (2.8 Å resolution). To ensure sufficient constraints on the phases from solvent flattening, the number of solvent pixels was not only scaled up in accordance with the grid spacing, but increased by a further 20%. To reduce series-termination errors and improve the phases for the high-resolution data, the phase-extension process was allowed to continue beyond the limit of the observed data, generating calculated amplitudes as well as phases. Cycles of averaging were continued until convergence (cycle 719). The final averaging R value [from (2)] was 14.9% and the correlation coefficient C [from (3)] was 0.92 on all data from infinity to 2.8 Å resolution. The accuracy of the phasing at higher resolution is indicated by the R factor and the correlation coefficient C at 3 Å, which are 24% and 0.79, respectively (Fig. 4*a*), and by the interpretability of the maps. The behaviour of the phase refinement is illustrated in Figs. 4(*a*) and (*b*).

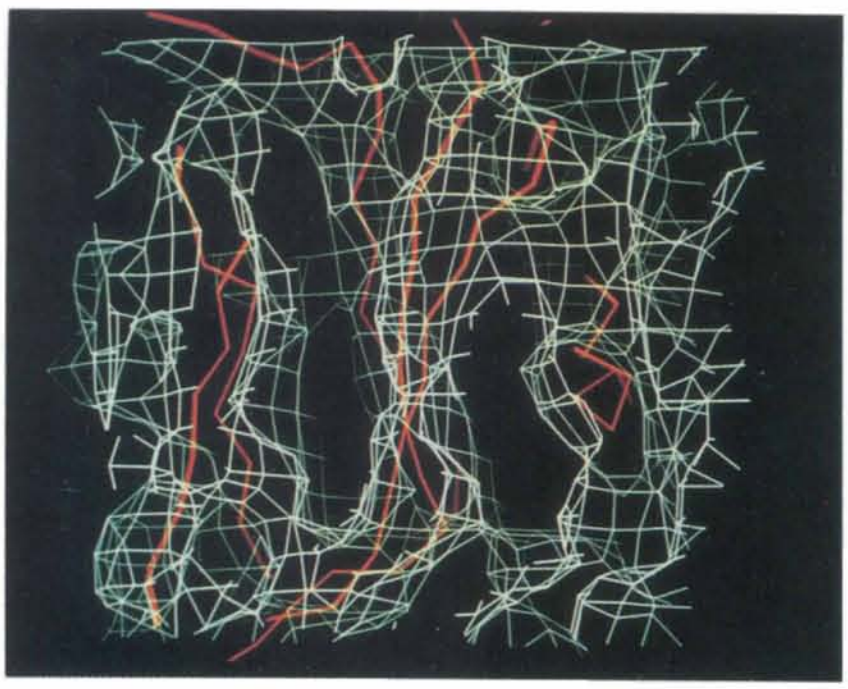
Electron-density maps were computed with amplitude coefficients of $2|F_{obs}|-|F_{averaged}|$ and $\alpha_{averaged}$ phases for viewing the icosahedral asymmetric unit.

A published sequence was available for the complete nucleotide sequence of the RNA coding for the primary translation product of foot-and-mouth disease virus strain O_1 BFS 1860 (Forss, Strebel, Beck, & Schaller, 1984). This had previously been aligned with other picornaviral sequences (Palmenberg, 1989).

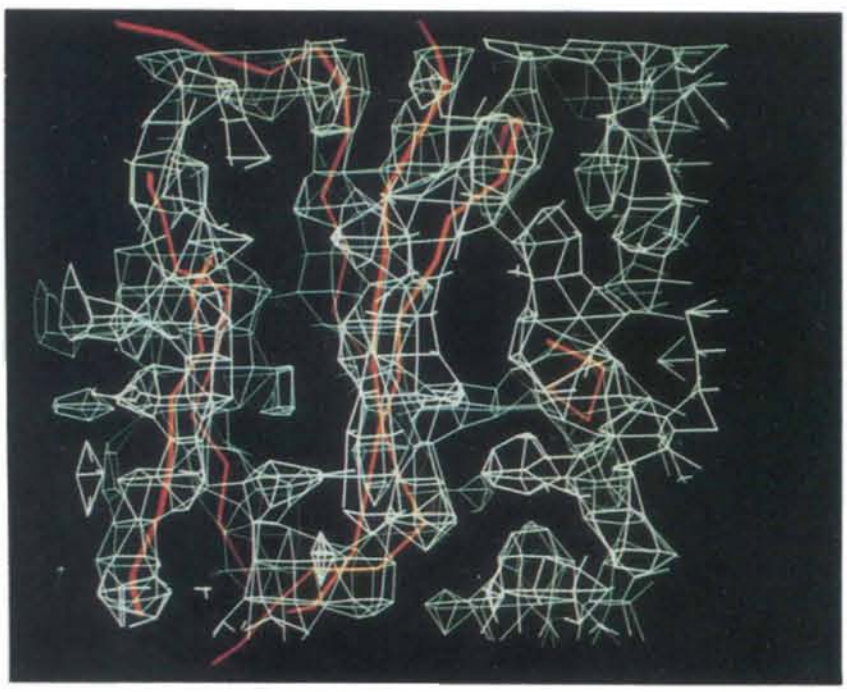
The main chain could be traced at 4.5 Å resolution. At 3.5 Å, side chains were identifiable, enabling sequence alignment by matching a characteristic stretch of residues with the electron-density features. A model was built at this resolution, fitting the sequence using appropriate stretches of the HRV14 coordinates according to the sequence alignment. This was achieved using an Evans and Sutherland PS390 graphics system and the *FRODO* program (Jones, 1978). No corrections to the published amino acid sequence for the capsid proteins appeared necessary. The initial model was improved using the 2.9 Å resolution map (see Fig. 4*b*) in which the carbonyl atoms were clearly identifiable, enabling the determination of the plane of the peptide. Many water molecules were also evident. Over 90% of the capsid residues were clearly visible in the electron-density map. In total, 6333 atoms were built into the 736 amino acid residues in the four distinct chains. The map was very easy to interpret. Aside from disordered terminii in VP2 and VP4, there were only two breaks in the main-chain density. One of these corresponded to an entire surface loop (the so-called G–H loop of VP1) (Acharya *et al.*, 1989). To check that the envelope-determination procedure had not erroneously truncated any surface features, the envelope was displayed superimposed on an unaveraged untruncated electron-density map calculated with $2|F_{obs}|-|F_{averaged}|$ amplitudes and $\alpha_{averaged}$ phases. The electron-density distribution in this map was extremely clear with no remarkable variation between the five copies. The electron density fitted comfortably within the envelope with no regions obviously impinging on its boundaries. Some diffuse density within the envelope was evident in this map, corresponding to the less well ordered parts of the structure. Later, an electron-density map was calculated with $|F_{obs}|-|F_{calc}|$ amplitudes and α_{calc} phases, where F_{calc} and α_{calc} were derived from the atomic model after least-squares refinement (for the protein only) and hence could not be affected by the envelope. This would clearly show any density truncated by the envelope; again no such density was observed.



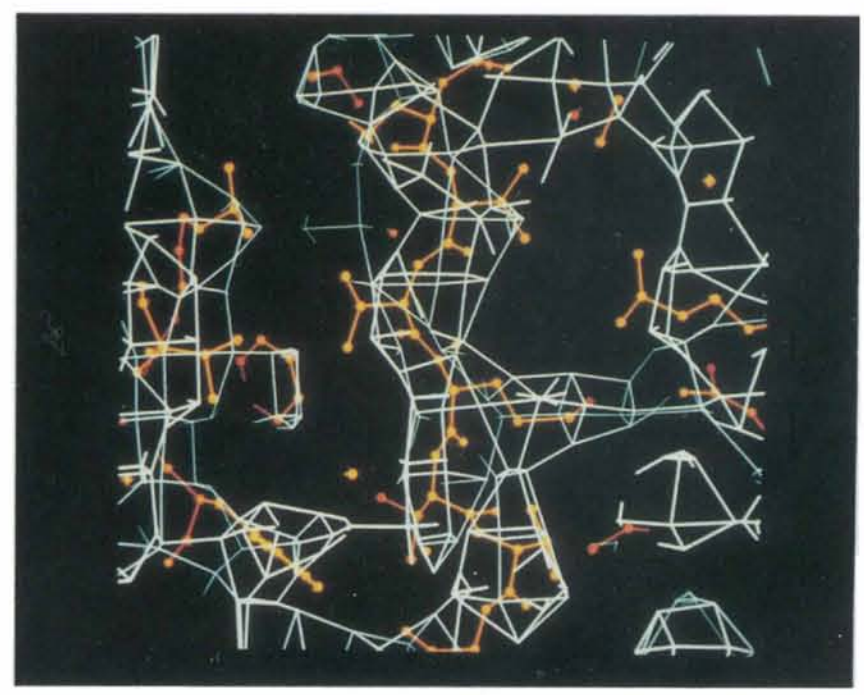
(*a*)



(i)



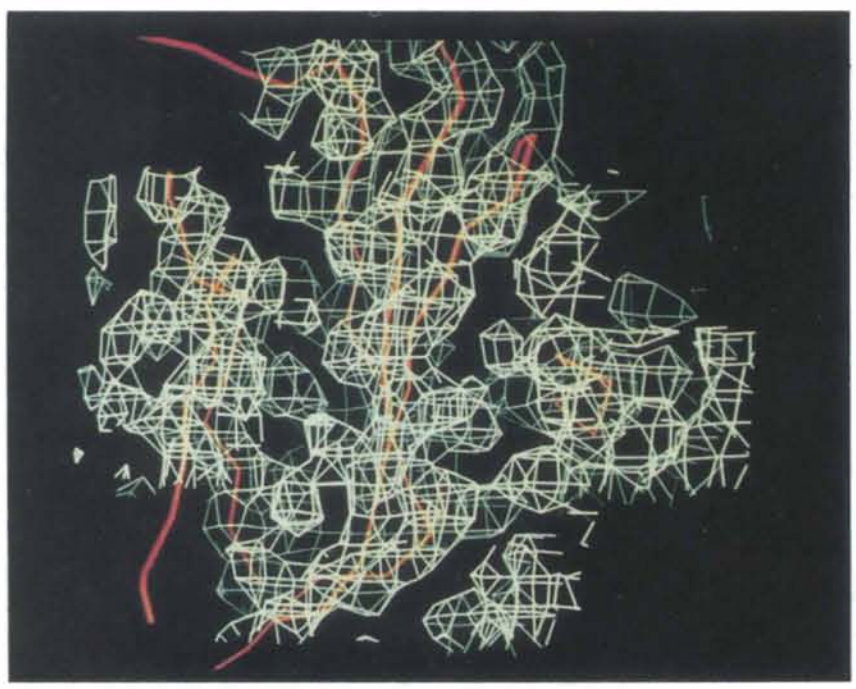
(ii)



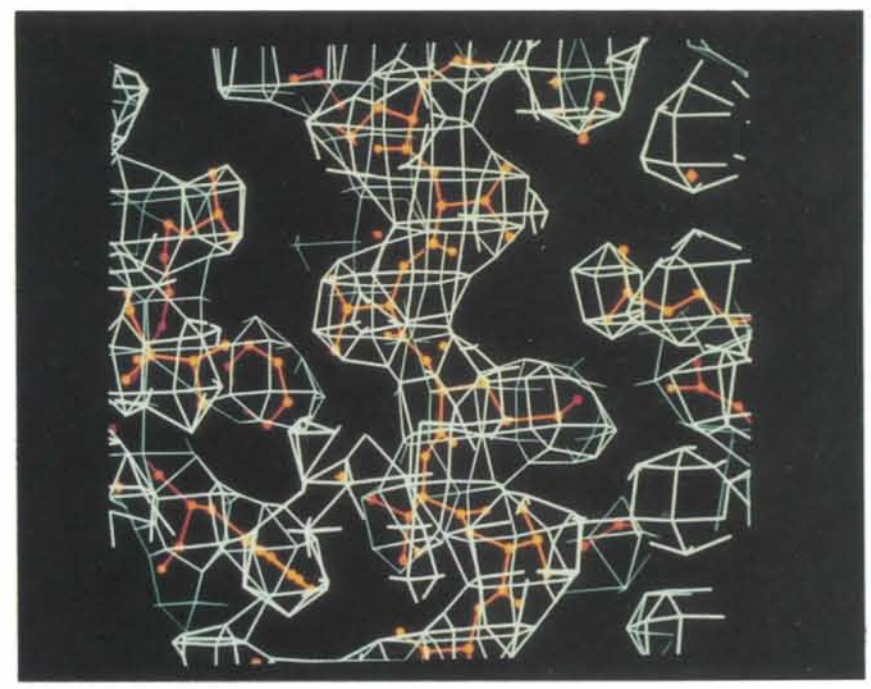
(iii)

(*b*)

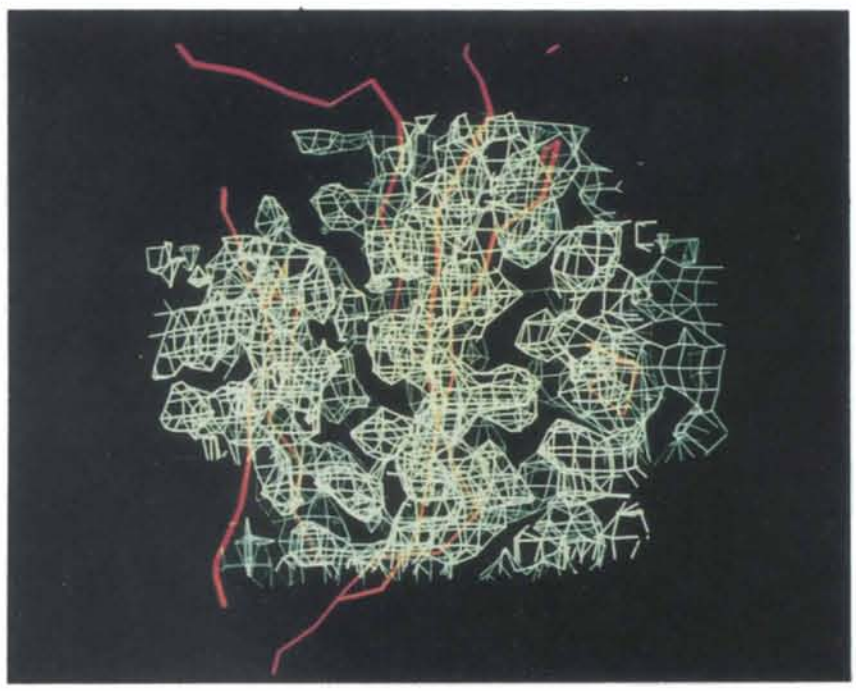
Fig. 4. Phase-extension summary. (*a*) *R* factor [(2)] and correlation coefficient *C* [(3)] plots for the phase extension from 8 to 2·8 Å (cycle 719). Cycle 0 is starting phases (unaveraged). (*b*) Identical portions of electron density maps (white), calculated at intervals in the phase extension, with coefficients $2(F_{obs} - F_{averaged})$, $\alpha_{averaged}$ and contoured at approximately one tenth of the maximum. (i) 6 Å (cycle 265), (ii) and (iii) 4·5 Å (cycle 528), (iv) and (v) 3·5 Å, (vi) and (vii) 2·9 Å [(iii), (v) and (vii) show an enlargement of the central portion of (ii), (iv) and (vi)]. Either the C_α backbone or an all-atom trace is superimposed in red.



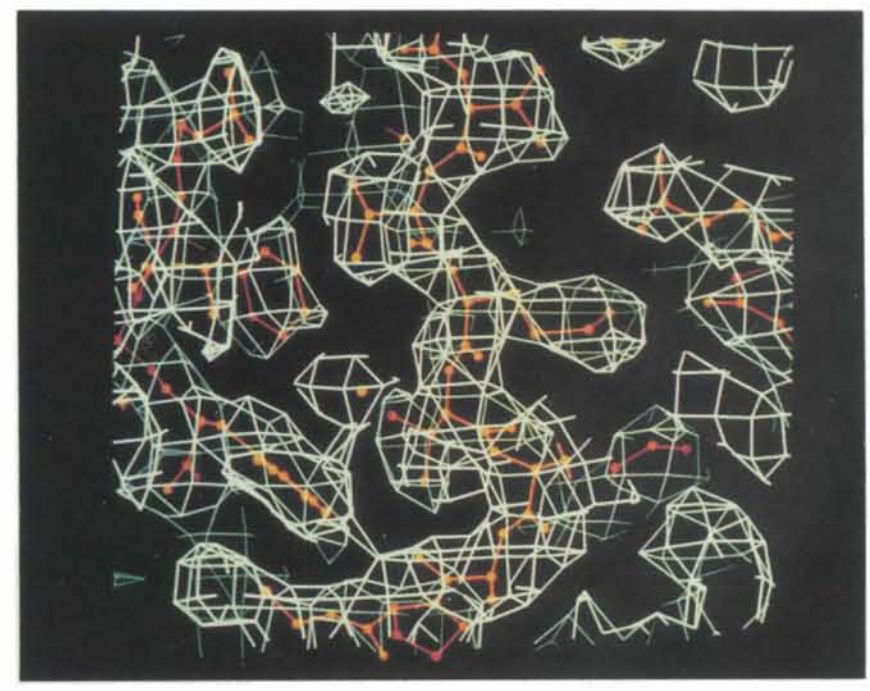
(iv)



(v)



(vi)



(vii)

Fig. 4. (*b*) (*cont.*)

Indeed, this map confirmed the original maps in all respects; peaks that had not been modelled as protein appeared clearly (these were mostly explicable as water molecules).

The *R* factor [from (2)] indicating the agreement of structure factors calculated from the model built into the map produced by phase extension with the observed amplitudes was 34.1% (including all data) (25.7% for those data with amplitudes greater than the mean amplitude for that resolution shell). The low value of the *R* factor reflects the clarity of the electron-density map, itself dependent on the quality of the phases (which were extremely accurate as a consequence of the symmetry of the viral capsid), the accuracy of the diffraction measurements and the inclusion of fragments from a stereochemically correct model in part of the building. Refinement (*X-PLOR*; Brünger, 1988) reduced the *R* factor on all data to 17% without manual intervention and without the inclusion in the model of bound water molecules. The r.m.s. deviation from ideal bond lengths was 0.017 Å (details of the refinement will be published elsewhere). Coordinates have been deposited with the Protein Data Bank (Bernstein *et al.*, 1977).*

We would like to thank Graham Fox for growing the virus crystals and David Rowlands, Fred Brown and David Phillips for their support and interest. In addition, we are indebted to all the staff at the SERC Synchrotron Radiation Facility at Daresbury and to those people from the Laboratory of Molecular Biophysics who helped with data-collection trips and computing. We are especially grateful to John Mann and David Goodrich for attendance at the synchrotron trips as disease security officers. The work was supported by the MRC.

* Atomic coordinates have been deposited with the Protein Data Bank, Brookhaven National Laboratory (Reference: 1BBT), and are available in machine-readable form from the Protein Data Bank at Brookhaven. The data have also been deposited with the British Library Document Supply Centre as Supplementary Publication No. SUP37069 (as microfiche). Free copies may be obtained through The Technical Editor, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England.

References

ACHARYA, R., FRY, E., STUART, D. I., FOX, G., ROWLANDS, D. & BROWN, F. (1989). *Nature (London)*, **337**, 709-716.

BERNSTEIN, F. C., KOETZLE, T. F., WILLIAMS, G. J. B., MEYER, E. F., BRICE, M. D., ROGERS, J. R., KENNARD, O., SHIMANOUCHI, Y. & TASUMI, M. (1977). *J. Mol. Biol.* **112**, 535-542.

BRICOGNE, G. (1976). *Acta Cryst.* A**32**, 832-847.

BRÜNGER, A. T. (1988). *X-PLOR* Version 1.3. *Crystallographic Computing* 4, edited by N. W. ISAACS & W. R. TAYLOR. pp. 126-140. Oxford Univ. Press.

CROWTHER, R. A. (1972). *The Molecular Replacement Method*, edited by M. G. ROSSMANN, pp. 173-178. New York: Gordon and Breach.

FORSS, K., STREBEL, K., BECK, E. & SCHALLER, H. (1984). *Nucleic Acids Res.* **12**, 6587-6601.

FOX, G., STUART, D., ACHARYA, K. R., FRY, E., ROWLANDS, D. & BROWN, F. (1987). *J. Mol. Biol.* **196**, 591-597.

FRENCH, S. & WILSON, K. (1978). *Acta Cryst.* A**34**, 517-525.

FRY, E., LOGAN, D. & STUART, D. (1992). In *Methods in Molecular Biology*, Vol. 7, *Physical Methods*, edited by B. MULLOY. In the press.

HOGLE, J. M., CHOW, M. & FILMAN, D. J. (1985). *Science*, **229**, 1358-1365.

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

KABSCH, W. (1988). *J. Appl. Cryst.* **21**, 67-71.

LIDDINGTON, R. C., YAN, Y., MOULAI, J., SAHLI, R., BENJAMIN, T. L. & HARRISON, S. C. (1991). *Nature (London)*, **354**, 278-284.

LUO, M., ROSSMANN, M. G. & PALMENBERG, A. C. (1988). *Virology*, **166**, 503-514.

LUO, M., VRIEND, G., KAMER, G., MINOR, L., 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., WILLINGMAN, P., ILAG, L. L., KRISHNASWAMY, S., ROSSMANN, M. G., OLSON, N. H., BAKER, T. S. & INCARDONA, N. L. (1992). *Nature (London)*, **355**, 137-143.

MORGUNOVA, E., FRY, E., MIKCHAILOV, A., VAINSHTEIN, B., DAUTER, Z., WILSON, K. & STUART, D. (1992). Submitted to *Nature (London)*.

OLSON, A. J., BRICOGNE, G. & HARRISON, S. C. (1983). *J. Mol. Biol.* **171**, 61-93.

PALMENBERG, A. C. (1989). *Molecular Aspects of Picornavirus Infection and Detection*, edited by B. L. SEMLER & E. EHRENFIELD, pp. 211-241. Washington, DC: American Society for Microbiology.

PARRY, N., FOX, G., ROWLANDS, D., BROWN, F., FRY, E., ACHARYA, R., LOGAN, D. & STUART, D. (1990). *Nature (London)*, **347**, 569-572.

RAYMENT, I. (1983). *Acta Cryst.* A**39**, 102-116.

ROSSMANN, M. G., ARNOLD, E., ERIKSON, J. W., FRANKENBERGER, E. A., GRIFFITH, J. P., JOHNSON, J. E., KAMER, G., LUO, M., MOSSER, A. C., 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.

SCHUTT, C. E. & WINKLER, F. K. (1977). *The Rotation Method in Crystallography*, edited by U. W. ARNDT & A. J. WONACOTT, pp. 173-186. Amsterdam: North-Holland.

STUART, D. I., LEVINE, M., MUIRHEAD, H. & STAMMERS, D. K. (1979). *J. Mol. Biol.* **134**, 109-142.

TANAKA, N. (1977). *Acta Cryst.* A**33**, 191-193.

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

VALEGÅRD, K., LILJAS, L., FRIDBORG, K. & UNGE, T. (1990). *Nature (London)*, **345**, 36-41.

WANG, B. C. (1985). *Methods Enzymol.* **115**, 90-112.

WILSON, K. S., STURA, E. A., WILD, D. L., TODD, R. J., STUART, D. I., BABU, Y. S., JENKINS, J. A., STANDING, T. S. & JOHNSON, L. N. (1983). *J. Appl. Cryst.* **16**, 28-41.