eventual solution is unique, but rather to explore the limits of deficiency in data and the initial phasing model to attain convergence to the correct solution.

Unlike the more complicated model used by Rayment (1983), a simple spherical-shell model was used for a virus. Similarly, a solid sphere might be used for a protein molecule with sufficiently high noncrystallographic symmetry. It was shown that the centricity of the phases could be eliminated automatically as resolution increased under the stringent conditions where the inclination between the molecular and crystallographic twofold axes was only 2.5°. It was also shown that the choice of a suitable outer radius is critically important. Although time-consuming, systematic trials using phase extension can be used to establish a suitable external radius; the internal radius is of less importance. A parallel study (Chapman, Tsao & Rossmann, 1992) shows how the radii can be determined directly from the diffraction data.

We are grateful to Jin-bi Dai, Jeffrey T. Bolin, Janet L. Smith, John E. Johnson and Craig Smith for useful discussions. We thank Sharon Wilder for typing the manuscript. The work was supported by grants from the National Institutes of Health and the National Science Foundation.

References

ABAD-ZAPATERO, C., ABDEL-MEGUID, S. S., JOHNSON, J. E., LESLIE, A. G. W., RAYMENT, I., ROSSMANN, M. G., SUCK, D. & TSUKIHARA, T. (1980). *Nature (London)*, **286**, 33-39.

ARGOS, P., FORD, G. C. & ROSSMANN, M. G. (1975). *Acta Cryst.* A**31**, 499-506.

ARNOLD, E. & ROSSMANN, M. G. (1986). *Proc. Natl Acad. Sci. USA*, **83**, 5489-5493.

ARNOLD, E., VRIEND, G., LUO, M., GRIFFITH, J. P., KAMER, G., ERICKSON, J. W., JOHNSON, J. E. & ROSSMANN, M. G. (1987). *Acta Cryst.* A**43**, 346-361.

BRICOGNE, G. (1974). *Acta Cryst.* A**30**, 395-405.

CHAPMAN, M. S., TSAO, J. & ROSSMANN, M. G. (1992). *Acta Cryst.* A**48**, 301-312.

GAYKEMA, W. P. J., HOL, W. G. J., VEREIJKEN, J. M., SOETER, N. M., BAK, H. J. & BEINTEMA, J. J. (1984). *Nature (London)*, **309**, 23-29.

HARRISON, S. C., OLSON, A. J., SCHUTT, C. E., WINKLER, F. K. & BRICOGNE, G. (1978). *Nature (London)*, **276**, 368-373.

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

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

JOHNSON, J. E., AKIMOTO, T., SUCK, D., RAYMENT, I. & ROSSMANN, M. G. (1976). *Virology*, **75**, 394-400.

LILJAS, L., UNGE, T., JONES, T. A., FRIDBORG, K., LÖVGREN, S., SKOGLUND, U. & STRANDBERG, B. (1982). *J. Mol. Biol.* **159**, 93-108.

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.

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. (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. & ERICKSON, J. W. (1983) *J. Appl. Cryst.* **16**, 629-636.

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

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

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

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

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

Acta Cryst. (1992). A**48**, 301-312

Ab Initio Phase Determination for Spherical Viruses: Parameter Determination for Spherical-Shell Models

BY MICHAEL S. CHAPMAN,* JUN TSAO† AND MICHAEL G. ROSSMANN

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

(*Received* 22 *July* 1991; *accepted* 5 *November* 1991)

Abstract

The structure determination of canine parvovirus depended on the extension of phases calculated initially from a spherical-shell model [Tsao, Chapman, Wu, Agbandje, Keller & Rossmann (1992). *Acta Cryst.* B**48**, 75-88]. Such *ab initio* phasing holds the promise of obviating initial experimental phasing by isomorphous or molecular replacement, thereby expediting the structure determinations of spherical virus capsids. In this paper, it is shown how parameters such as radii, DNA density and particle

* To whom correspondence should be addressed.

† Current address: Department of Biochemistry, Center for Macromolecular Crystallography, 262BHS, THT 79, University of Alabama, University Station, Birmingham, Alabama 35294, USA.

0108-7673/92/030301-12$03.00

© 1992 International Union of Crystallography

positions may be determined and refined from diffraction data with sufficient precision to start phase extension from 20 Å resolution for a virus of approximately 122 Å radius.

Introduction

The capsids of spherical viruses have long been the object of intensive structural investigations, both because of their intrinsic biological interest and because the atomic structure provides understanding of viral infection, immune response and drug design (reviewed in Liljas, 1986; Rossmann & Johnson, 1989; Harrison, 1990; Chapman, Giranda & Rossmann, 1991). Largely because of their size, usually exceeding a diameter of 190 Å, their crystallographic structure determination remains challenging and relatively few have been solved. Phases can be refined from within a large convergence radius to considerable accuracy by imposing highly redundant icosahedral noncrystallographic symmetry (Rossmann, 1990). A major bottleneck remains the determination of initial phases. In the past, initial phases have been determined by isomorphous replacement, for example, tomato bushy stunt virus (Harrison, Olson, Schutt, Winkler & Bricogne, 1978), satellite tobacco necrosis virus (Liljas, Unge, Jones, Fridborg, Lövgren, Skoglund & Strandberg, 1982), human rhinovirus 14 (Rossmann, Arnold, Erickson, Frankenberger, Griffith, Hecht, Johnson, Kamer, Luo, Mosser, Rueckert, Sherry & Vriend, 1985), poliovirus (Hogle, Chow & Filman, 1985) and black beetle virus (Hosur, Schmidt, Tucker, Johnson, Gallagher, Selling & Rueckert, 1987). It has also been possible to solve, with homologous atomic models, the structures of related viruses by molecular replacement, for example, mengovirus (Luo, Vriend, Kamer, Minor, Arnold, Rossmann, Boege, Scraba, Duke & Palmenberg, 1987), human rhinovirus 1A (Kim, Smith, Chapman, Rossmann, Pevear, Dutko, Felock, Diana & McKinlay, 1989) and foot-and-mouth disease virus (Acharya, Fry, Stuart, Fox, Rowlands & Brown, 1989). In some cases phase refinement was started at resolutions as low as 8 Å (Luo *et al.*, 1987). It has also been shown that completely unrelated models can be used, provided that phase refinement is initiated at low enough resolution (Valegård, Liljas, Fridborg & Unge, 1990; McKenna, Xia, Willingmann, Ilag, Krishnaswamy, Rossmann, Olson, Baker & Incardona, 1991). This progression suggested that it might be possible to extend phases from low resolution after calculating initial phases from a spherical shell of uniform density.

Canine parvovirus (CPV) is a small non-enveloped single-stranded DNA virus. Parvoviruses infect the cells of proliferating tissues, causing serious diseases that can be fatal in many mammals, including man (reviewed by Pattison, 1990). The structure of CPV was of interest, not only because of its potential use in medical research, but also to see whether the β-barrel motif, common to picornaviruses and many plant and insect RNA viruses (reviewed by Rossmann & Johnson, 1989; Harrison, 1990), would be found in other families of viruses. The CPV structure was the first of its family to be determined, so it was not possible to determine initial phases from the structure of a related virus. Crystal production was difficult and it took two years to complete native-data collection. Therefore, structure determination by conventional isomorphous replacement methods appeared to be unrealistic and unattractive because it would have required the collection of near-complete data sets of several potential derivatives. To avoid this, attempts were initiated to determine phases *ab initio* as discussed by Tsao, Chapman & Rossmann (1992) and in this paper.

A similar approach had been used to determine phases for southern bean mosaic virus (SBMV) (Johnson, Akimoto, Suck, Rayment & Rossmann, 1976) and polyoma virus (Rayment, Baker, Caspar & Murakami, 1982) to 22.5 Å resolution. With SBMV, initial phases were calculated from the Fourier transform of a solid sphere with a diameter of 281 Å, the nearest-neighbor contact distance within the crystal. The Fourier transform agreed well with spherically averaged structure-factor amplitudes between 60 and 35 Å resolution. Phases were refined through the imposition of icosahedral symmetry and extended to 22.5 Å resolution where the agreement between observed and back-transformed structure-factor amplitudes was nearly random. In SBMV, the particle center lies in a special position with 32 symmetry, limiting the resolution to phases that could be extended, since icosahedral averaging could never break the centric nature of the initial phases. With polyoma virus (Rayment, Baker & Caspar, 1983), phases were initially calculated from a decorated spherical-shell model based on electron microscopy results. This broke the center of symmetry and permitted successful phase extension to 22.5 Å resolution.

In attempts to avoid isomorphous replacement, efforts have been directed to initiate phase extension for several viral structures using a variety of spherical and other low-resolution phasing models (Tsao, 1990; McKenna *et al.*, 1991). Prior to the structure determination of φX174, about 15 unsuccessful attempts were made to extend phases from various models, each requiring extensive use of supercomputer time in phase refinement before determining that the initial phases were inadequate. Simulations with structure factors calculated from a model (Tsao, Chapman & Rossmann, 1992) show that the model outer radius and position are critical to the success of phase extension. An error of 2% in outer radius was sufficient to introduce significant error in the phases, while 3% was sufficient to completely foil phase

refinement. (Determination of the external radius of CPV from packing considerations would have resulted in an error of 5%.) Similarly, displacement of the particle position by 0.006 fractional units introduced a mean phase error after refinement of 30° at 20 Å resolution. Thus, the motivation was strong to develop a method capable of determining spherical-model parameters directly from the diffraction data.

Earlier treatments of low-resolution diffraction from viruses considered the Fourier transform (in spherical polar coordinates) as a series of radially dependent Bessel functions modulated by directionally dependent icosahedral harmonics (Finch & Holmes, 1967). These treatments show that the non-spherically symmetric topology of the virus is expected to give a significant component to the structure amplitudes. The non-spherical modulations were exploited by Jack & Harrison (1975) in structural interpretation of the solution scattering of cucumber mosaic and polyoma viruses and were cited as the reason for increasing loss of spherical symmetry in the diffraction pattern of tomato bushy stunt virus between 100 and 40 Å resolution (Harrison, 1969). Although a low-order non-spherical approximation might be useful in phase initiation where a low-resolution spherically symmetric model is inadequate, a simple spherical shell was preferred in this work, to minimize the chances of compromising phase refinement through possible errors in the initial phasing model. Non-spherical modulations should not introduce systematic error in spherical parameters, because the spherically averaged value of the structure amplitude, as given in equation (18) of Finch & Holmes (1967), is independent of non-spherical modulations. In using a spherical shell, the non-spherical modulations pose two other potential problems.

(1) Spherical parameters must be fitted to the spherical component of the structure amplitudes, which become smaller as the resolution is increased. This paper shows that such fitting is possible even at resolutions where the spherical component is minor, given that there is a data-to-parameter ratio of about 200.

(2) The initial phase of a reflection may be based on only a small component of the structure factor, requiring that the phases be refined from near-random initial values.

The methods discussed here were developed during the structure determination of CPV. Spherical-shell parameters, determined through an early application of these methods (when only 35% of the data were available), were used to calculate the initial phases for CPV. These phases were sufficient for extension to 9 Å and to determine the site of substitution of a single isomorphous derivative (SIR) (Tsao, Chapman, Wu, Agbandje, Keller & Rossmann, 1992). SIR phases were used as a starting point for extension to 3.25 Å, although Tsao, Chapman, Wu *et al.* (1992) showed that, if the center of icosahedral symmetry had been refined during the extension from 20 to 9 Å, it would not have been necessary to use SIR phase information. During the phase extension, the spherical-shell parameters were refined again, using a more complete data set. It is these improved parameters (external and internal radii, particle position and now also the relative nucleic acid and protein densities) that are presented here, because, as will be demonstrated, they would have generated more accurate initial phases if it had proved necessary to restart the phase extension. Related papers investigate how much error in data and model parameters may be tolerated (Tsao, Chapman & Rossmann, 1992), describe the structure determination of CPV (Tsao, Chapman, Wu *et al.*, 1992) and the results (Tsao, Chapman, Agbandje, Keller, Smith, Wu, Luo, Smith, Rossmann, Compans & Parrish, 1991).

Experimental methods

The virus was prepared according to the procedure of Paradiso (1981) and modified as in Luo, Tsao, Rossmann, Basak & Compans (1988) and Tsao *et al.* (1991). Solution scattering patterns were collected by methods similar to those used by Schmidt, Johnson & Phillips (1983), except that the sample concentration was about 10 mg ml^{-1} rather than 150 mg ml^{-1} because of the difficulty of propagating sufficient virus in mammalian cell lines. To compensate, long exposures of up to one week were required with a wider beam than had been used previously (Schmidt *et al.*, 1983) at a generator power of 35 kV × 25 mA. Cu $K\beta$ radiation was removed with a 0.01 mm thick nickel filter. Four sets of slits were used along a 2.40 m helium path to collimate the beam. The sample-to-film distance was 268 mm.

Films were scanned with a 50 μm raster. Prior to radial averaging, the film center was estimated from the center of mass of optical density and accurately determined by maximizing $\sum|\Delta I/\Delta d^*|$, where I is the radially averaged film density, d^* is reciprocal resolution and the derivative is estimated numerically. The three films within each film pack were scaled to each other using a cubic function of optical density with coefficients fitted by linear least-squares refinement. A variance-weighted mean of the radially averaged optical densities (OD) from all films was calculated assuming Poisson counting statistics for each pixel. Programs to refine model parameters were tested by using data extracted from Fig. 3 of Wobbe, Mitra & Ramakrishnan (1984) and yielded radii of 124.0 and 75.6 Å for Kilham rat virus compared to the published values of 124.3 and 74.0 Å.

The data collection by oscillation photography and processing of the crystallographic data are described in Tsao, Chapman, Wu *et al.* (1992).

Theory

Optimization of model parameters was achieved by searches and then least-squares refinement in reciprocal space against the observed X-ray data. For a set of spheres of uniform protein density ρ_P at positions $\mathbf{S}_n$ ($n = 1$ to N) in the unit cell, the structure factor $\mathbf{F}_s(\mathbf{h})$ for reflection $\mathbf{h}$ is

$$\mathbf{F}_s(\mathbf{h}) = \sum_{n=1}^{N} (U_0/V)\rho_P G_{\mathbf{h}}(R_0) \exp(2\pi j \mathbf{h}\cdot\mathbf{S}_n), \quad (1)$$

where U_0 is the volume of each sphere of radius R_0, V is the volume of the unit cell and $G_{\mathbf{h}}(R_0)$ is the Fourier transform of a solid sphere normalized to 1 at $|\mathbf{h}| = 0$ (Rossmann & Blow, 1962);

$$G_{\mathbf{h}}(R_0) = 3[\sin(2\pi|\mathbf{h}|R_0) - 2\pi|\mathbf{h}|R_0 \cos(2\pi|\mathbf{h}|R_0)]/(2\pi|\mathbf{h}|R_0)^3. \quad (2)$$

Similarly, for spherical shells with internal radius R_i, filled with nucleic acid of density ρ_{NA},

$$\mathbf{F}_s(\mathbf{h}) = \sum_{n=1}^{N} [(U_0/V)\rho_P G_{\mathbf{h}}(R_0) - (U_i/V) \times (\rho_P - \rho_{NA}) G_{\mathbf{h}}(R_i)] \exp(2\pi j \mathbf{h}\cdot\mathbf{S}_n) \quad (3)$$

$$= (\rho_P/V)\tfrac{4}{3}\pi R_0^3 \sum_{n=1}^{N} [G_{\mathbf{h}}(R_0) - (R_i^3/R_0^3) \times (1 - \rho'_{NA}) G_{\mathbf{h}}(R_i) \exp(2\pi j \mathbf{h}\cdot\mathbf{S}_n)], \quad (4)$$

where ρ'_{NA} is a relative density of the nucleic acid, defined as $\rho'_{NA} = (\rho_{NA}/\rho_P)$.

As the resolution increases, a real virus deviates more noticeably from that of a spherical shell. Let $\mathbf{F}_u(\mathbf{h})$ be the unknown non-spherical part of the structure factor $\mathbf{F}_{calc}(\mathbf{h})$, while $\mathbf{F}_s(\mathbf{h})$ is the spherical part represented by (2). Thus

$$\mathbf{F}_{calc}(\mathbf{h}) = \mathbf{F}_s(\mathbf{h}) + \mathbf{F}_u(\mathbf{h}) \quad (5)$$

and

$$\mathbf{F}_{calc}^2 = \mathbf{F}_s^2 + \mathbf{F}_u^2 - |\mathbf{F}_s||\mathbf{F}_u| \cos(\Delta\alpha), \quad (6)$$

where $\Delta\alpha$ is the difference in phase between $\mathbf{F}_s$ and $\mathbf{F}_u$. For the average of a sufficiently large number of reflections, it can be assumed that $\Delta\alpha$ is random and, hence,

$$\mathbf{F}_{calc}^2 \simeq \mathbf{F}_s^2 + \mathbf{F}_u^2. \quad (7)$$

Let $\mathbf{F}_u^2$ be approximated by a smoothly increasing function with resolution:

$$\mathbf{F}_u^2 = [J + K \exp(L|\mathbf{h}|) + M \exp(N|h|^2)]^2. \quad (8)$$

This is not intended to approximate closely the non-spherical component of individual amplitudes, which is likely to show considerable variation (Finch & Holmes, 1967; Harrison, 1969). However, the approximation facilitates the fitting of a spherical-shell transform that has zero amplitude at nodal resolutions to the observed structure amplitudes.

Using (7), (4), (2) and (8) it is now possible to set up a least-squares refinement procedure to minimize $\sum_h (\mathbf{F}_{obs}^2 - k\mathbf{F}_{calc}^2)^2$ from which the parameters R_0, R_i, ρ'_{NA}, J, K, L and M, the scale factor k and the positions $\mathbf{S}_n$ can be refined. The latter will be related by space-group symmetry and, hence, it is in general only necessary to refine $\mathbf{S}_1$, the particle position in the reference asymmetric unit.

Once the model parameters have been determined, (4) can be used to calculate phases for initial phase extension. The Fourier coefficient used to represent the electron density for averaging can be weighted by figures of merit, m. These can be estimated by the ratio of $\mathbf{F}_s$ and $\mathbf{F}_{calc}$.

The normal least-squares equations that were used to calculate the shifts in the model parameters during their refinement were sometimes not well conditioned. This is because some of the parameters may be correlated with one another. For example, when the internal density, ρ'_{NA}, is near unity, internal radius $\mathbf{R}_i$ has little effect on the overall Fourier transform of the model. To avoid large correlated changes of parameters that have little effect on the quality of the fit, in the initial cycles of refinement, eigenvalue filtering was used to accept only those shift vectors that gave greatest improvement of the residual (Diamond, 1971).

With most space groups there are certain 'special' positions of a spherical object which would give zero calculated intensity for a class of reflections. For example, in space group $P2_1$, a spherical object placed at $(\frac{1}{4}, y, \frac{1}{4})$ would give zero intensities for all reflections with $h + k + l =$ odd. Weak and 'non-positive' intensities are usually rejected during processing, leaving the accepted intensities with experimental errors that tend to *increase* their size. This effect would increase the apparent distance of the particle away from the 'special' position. Hence, the class of reflections for which the calculated intensities are small or zero may have to be omitted from the data set.

As resolution is increased, (2) quickly approximates to its high-resolution limit,

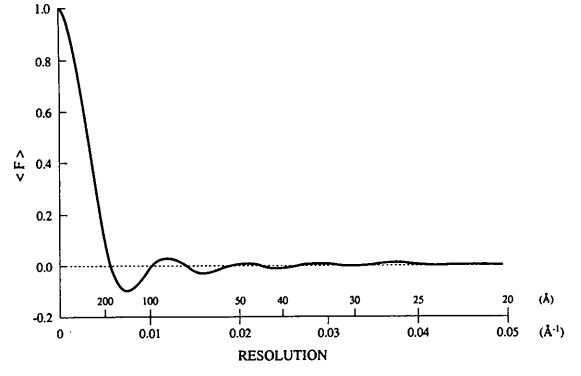
$$G = 3\cos(2\pi|\mathbf{h}|R)/(2\pi|\mathbf{h}|R)^3, \quad (9)$$

which is a damped cosine oscillation. For a virus with outer radius of 120 Å, a series of models with outer radii differing by 12 Å will have approximately the same calculated intensity distribution near 20 Å resolution (Fig. 1). The difference between radii, ΔR, that give similar intensity distributions is inversely proportional to both radius and $|\mathbf{h}|$. Unless special efforts have been made to collect exclusively low-resolution data, this ambiguity can be difficult to resolve.

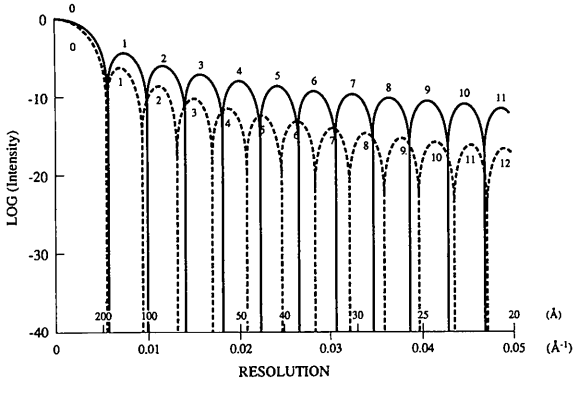
In the absence of very low resolution crystallographic data there needs to be an independent estimate of the outer radius, of accuracy better than half ΔR, to determine the absolute model radii. X-ray solution scattering is one of several methods that may

be applicable. With minor modification, the same refinement methods can be used to optimize model parameters to solution scattering data. The measured intensity is dependent only on resolution, while position is irrelevant to the calculation of intensities. The radii and ρ'_{NA} can be defined similarly to the crystallographic data. The background can be approximated in the same way as $|\mathbf{F}_u|$ [(8)].

If the observed intensities are fitted with I_c's that are one node shifted from I_c's calculated with the correct model, then G will be of inverted sign and calculated phases will be in error by exactly 180° [(2)]. This is the Babinet opposite solution and maps calculated with such phases would have the negative of the correct electron density. Note, however, that the periodicity of (9) is dependent on the radius and, hence, the Babinet opposite alternative solution can be expected to be valid only over a limited resolution range.



(*a*)



(*b*)

Fig. 1. Fourier transforms of spheres for different radii. (*a*) The Fourier transform of a sphere of radius 120 Å, plotted as a function of resolution. Note how it rapidly becomes a damped oscillating function (*b*). In a diffraction experiment, intensities not structure factors are measured. Plots show the log of the intensity for radii of 120 Å (solid line) and 132 Å (dashed line). Between 25 and 20 Å resolution, calculated intensities are similar, although the peaks are shifted by one node relative to each other. Indeed, if the low-resolution data are not measured, there may be a series of discretely different radii (*e.g.* 108, 120, 132, 144, ...) that fit the data equally well.

Refinement of model parameters for CPV

Unfortunately, due to the low sample concentration, long exposure and wide incident beam, the solution scattering of CPV exhibited more peak blurring and lower signal-to-background ratio (Fig. 2) than has been obtained for plant viruses (Schmidt *et al.*, 1983). The initial background coefficients were determined by fitting an exponential function to the observed optical density at the nodal points. Model and background parameters were then refined simultaneously by least squares from many different starting points. Residual errors between calculated and observed optical densities of less than 4% were achieved with outer radii of about 128 Å, inner radii of either 89 or 75 Å and nucleic acid density (ρ'_{NA}) of 1.67. The outer radius from solution scattering agreed with crystal-packing calculations. The value of ρ'_{NA} agreed well with that found for cowpea mosaic virus (1.57–1.65) (Schmidt *et al.*, 1983). However, the precision of determining the outer radius from solution scattering, which extended to about 75 Å resolution, was insufficient to be certain of obtaining the correct Babinet solution for the X-ray data, which started at about 30 Å resolution. This would not have been a problem had it been possible to collect data of the quality of Schmidt *et al.* (1983).

The model parameters were then optimized against crystallographic data that were 85% complete between 30 and 13.5 Å resolution. The high resolution limit was selected because oscillations of a somewhat regular periodic nature in $\langle|\mathbf{F}_{obs}|\rangle$ appeared to extend to 13.5 Å resolution. In retrospect, this was excessively optimistic. Initially, refinement was against only the $h+k+l=$ even reflections because the particle was near a 'special' position. A smoothly rising approximation to $\langle|\mathbf{F}_u|\rangle$ was determined by fitting the

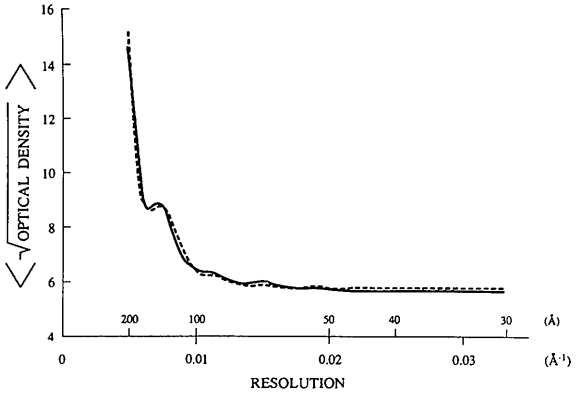


Fig. 2. Solution scattering. Observed optical densities (solid line) are compared to those calculated from a model with outer radius of 128.3 Å and inner radius of 89.4 Å (dotted line).

coefficients of (8) to the minima of the $\langle|F_0|\rangle$ distribution. Simultaneous refinement of these coefficients and model parameters proved unstable, so $\langle|F_u|\rangle$ coefficients were fixed. Parameter determination started with a two-dimensional R-factor search for the inner and outer radii, using a particle position of $(\frac{1}{4}, 0, \frac{1}{4})$ and ρ'_{NA} of 1.67 (Fig. 3). Physically reasonable solutions were found with outer radii of 130 and 120 Å and inner radii of 94 and 83 Å, respectively. For a partial specific volume of 0.72 $cm^3\ g^{-1}$, typical

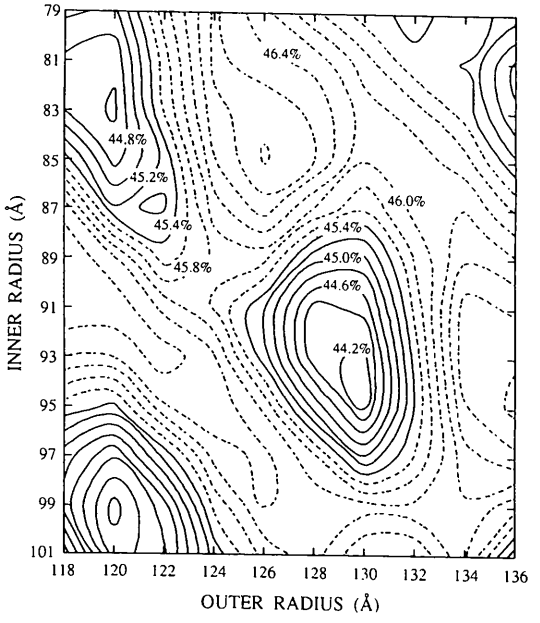


Fig. 3. R factor on F is plotted as a function of inner and outer radii in an initial search during which the position was set to $(\frac{1}{4}, 0, \frac{1}{4})$. Note that R factors are calculated for functions that include an approximation for $\langle|F_u|\rangle$ (unknown structure) and are, therefore, lower than the R factors for simple spherical-shell models.

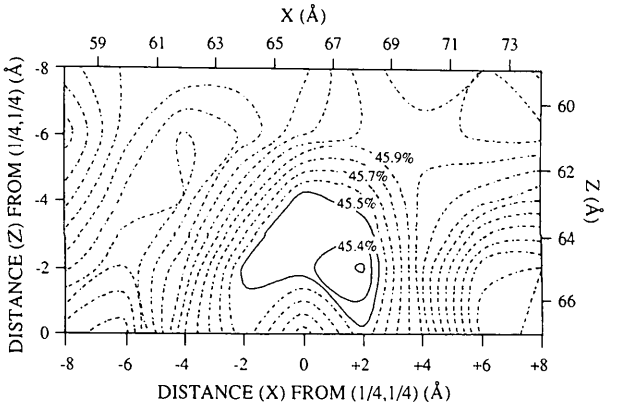


Fig. 4. An example of an R-factor search for the position of the particle center. This search was done with an outer radius of 120 Å and an inner radius of 83 Å using all data between 30 and 13.5 Å (Table 1). It is unnecessary to search $z > \frac{1}{4}$ because there is a twofold axis of symmetry at $(\frac{1}{4}, 0, \frac{1}{4})$ when the search is done with a centrosymmetric probe.

Table 1. *Searches and refinements to determine the position of the particle center*

		(x, z) displacement from $(\frac{1}{4}, 0, \frac{1}{4})$ (Å)	
Radii	Method	$h+k+l$ = even	All reflections
120/83	Search	(−1.5, −1.5)	(−1.8, −1.9)
120/83	Refinement	Unknown	(−0.9, −3.1)
130/94	Search	(−0.4, −1.3)	(0.3, −2.1)
130/94	Refinement	(−0.1, 0.0)	(1.0, −2.6)
	Refined at 3.8 Å*		(0.8, −0.8)

* The position was refined after phase extension by searching for the particle center that gave minimum root-mean-square deviation of the electron density when comparing the 60 asymmetric units related by noncrystallographic symmetry (Tsao, Chapman, Wu *et al.*, 1992).

of whole viruses (Table 6 of Smith, 1968), the outer radius would be approximately 118 Å. Assuming packing of spherical viruses within the crystal, the outer radius would be 129 Å. Previous structural investigations of spherical viruses showed that the capsid thickness is typically 30–40 Å (Schmidt *et al.*, 1983).

Both sets of radii were used to determine the position with an R-factor search on a 2 Å grid centered at $(\frac{1}{4}, y, \frac{1}{4})$. At each position the radii were fixed but a scale factor was refined. Searches were performed using all reflections and the $h+k+l$ = even reflections only (Fig. 4; Table 1). Retrospective comparison with the position refined after phase extension to 3.8 Å (Tsao, Chapman, Wu *et al.*, 1992) showed that the worst of these searches was in error by 2.4 Å.

The radii, position and scale factor were then simultaneously refined by least squares. The number of eigenvectors accepted was incremented from one to four. Alternative refinements were started with both possible radii (130 and 120 Å) and associated parameters. Refinement against the full data set, using only the $h+k+l$ = even reflections, gave rise to positions that differed by only 0.4 Å, suggesting that the particle center was sufficiently displaced from $(\frac{1}{4}, y, \frac{1}{4})$ that the $h+k+l$ = odd reflections need not be excluded any longer. It proved beneficial to refine occasionally just the scale factor, probably because the number of selected eigenvectors was less than the number of parameters. The DNA density was then allowed to change. The number of eigenvectors accepted was incremented from three to six with ten cycles of refinement for each. The overall drop in R factor was small, 0.53 and 0.71% for the models with an outer radii of 130.0 and 121.5 Å, respectively (Table 2). Noteworthy is the large drop in DNA density from 1.7 to about 1.2, probably because of the inclusion of relatively high resolution (13.5 Å) data where the effect of the DNA density would be attenuated because of higher disorder than in the protein.

Unfortunately, the small difference in R factor between the models with outer radii of 121.5 and

Table 2. *Refined parameters for models with outer radii of approximately* 130 *and* 120 Å *for data between* 30 *and* 13.5 Å *resolution*

Parameter	Refined values	
Outer radius (Å)	121.50	130.00
Inner radius (Å)	81.70	91.70
Nucleic acid density, ρ'_{NA}	1.16	1.32
x displacement from $(\frac{1}{4}, 0, \frac{1}{4})$	0.91	0.05
z displacement from $(\frac{1}{4}, 0, \frac{1}{4})$	−3.15	−2.61
Scale factor	88 139	93 619
R factor (%)	44.6	44.4

130.0 Å gave no indication of which was correct. Crystal packing and the solution scattering favored 130.0 Å, but comparison with the models for Kilham rat virus (outer radius = 120.6–124.3 Å) (Wobbe *et al.*, 1984) favored 121.5 Å. Phases calculated for models similar to both of these had, by now, been successfully extended to 13 Å (Tsao, Chapman, Wu *et al.*, 1992), converging on either the true solution or its Babinet inverse. This showed that either solution was sufficient to initiate phase extension towards atomic resolution. The structure determination enabled a retrospective assessment of the accuracy of starting phases calculated from the refined spherical-shell models.

Retrospective determination of the radii

(1) *From the atomic structure*

The radial distribution, $\varphi_{atom}(R)$, of atoms from the particle center (Fig. 5) shows that the inner and outer radii are approximately 86 and 124 Å, respectively. However, due to large surface features on CPV (Tsao *et al.*, 1991), the distribution is not a step function, as a spherical shell would imply. Therefore, the relevant radial cut-offs were not clearly determined. This problem can be solved by taking the Fourier transform of the distribution, and then determining the radii of the G function that fits optimally. Surprisingly, the optimal outer radius decreased as a function of the resolution (Table 3). For data in the resolution range of ∞ to 50 Å, the optimal radius was 128 Å, but for data in the 22 to 18 Å range it was 119 Å. When resoutions below 33 Å were omitted, an outer radius of 130 Å was only of slightly worse fit than 123 Å, but graphical comparison of the G functions (Fig. 6) showed that, by 22 Å, the G function for $R_0 = 130$ Å had inserted an extra node relative to a G function for $R_0 = 121.5$ Å.

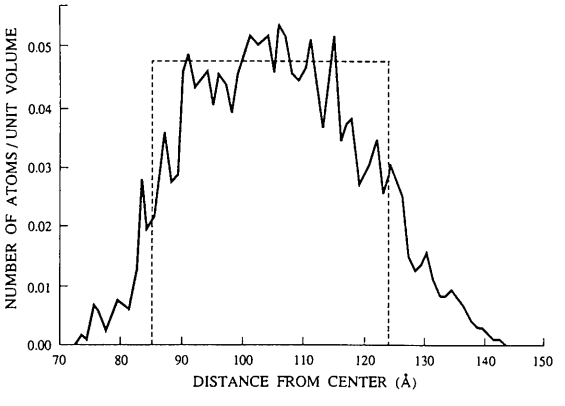


Fig. 5. Radial atomic density distribution for CPV. The distribution as calculated from the atomic coordinates is shown by a solid line. The dashed line is a step function of approximately equal area, fitted by eye to the distribution. This suggests that the best uniform-density shell model would have radii of about 86 and 124 Å.

Table 3. *Refinement of parameters against the Fourier transform of the spherically averaged structure*

Parameter	Refined values				
Resolution range (Å)	∞→50	70→40	55→25	25→20	22→18
Outer radius (Å)	128.4	125.8	123.4	123.3	119.3
Inner radius (Å)	82.8	84.2	87.7	78.0	78.9
R factor (%)	1.0	12.3	50.9	29.7	37.3

Together, these results show that 121.5 and 81.7 Å were reasonable model radii for phase extension from 20 Å. They also show that model parameters cannot be extrapolated from very low resolution, rather they should be refined using the data from which phases are to be extended. The resolution dependence for the determination of the outer radius arises presumably because CPV is not a spherical shell of uniform density.

(2) *From refined phases*

In the CPV structure determination (Tsao, Chapman, Wu *et al.*, 1992), phases from a spherical shell

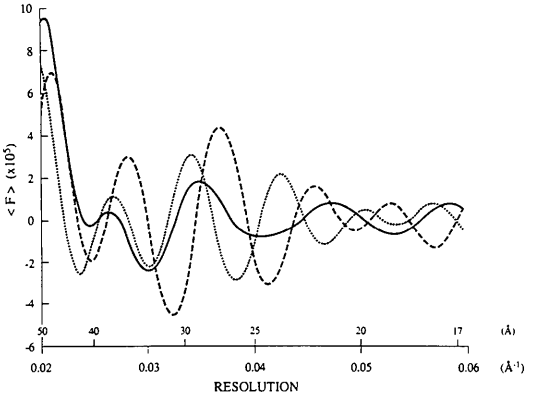
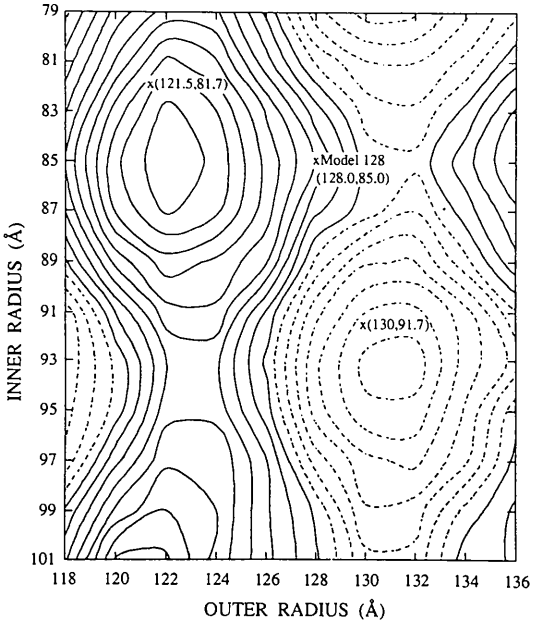


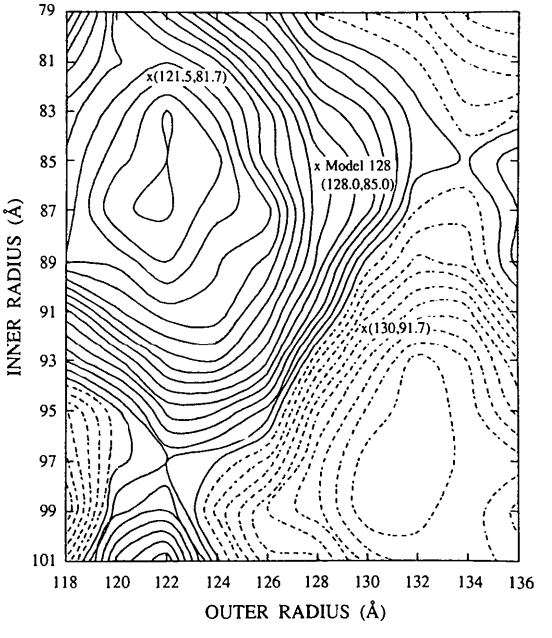
Fig. 6. The Fourier transform of the spherically averaged CPV atomic structure is shown as a solid line. The unscaled Fourier transforms are shown of spherical shells of radii 121.5 and 81.7 Å (dashed line) and 130.0 and 91.7 Å (dotted line), both of which agree well with that of the spherically averaged structure at resolutions lower than 50 Å resolution (not shown). At 22 Å resolution the transform with an outer radius of 130 Å is out of phase with both that of 121.5 Å and that of the spherically averaged structure. An outer radius of 121.5 Å is a compromise between optimal agreements at 30 and 20 Å resolutions.

had been extended to 9 Å resolution where they were used to determine the location of a heavy-atom derivative. Phases calculated from the isomorphous derivative between 15 and 9 Å resolution were refined and extended back to 150 Å resolution. Hence, these low-resolution phases are entirely independent of a starting spherical model and in this analysis are considered as the 'true' phases. During the phase back-extension the nucleic acid density within 70 Å of the particle center was repeatedly set to zero, as also was the external solvent. Hence, the 'true' phases will be biased by this assumption.

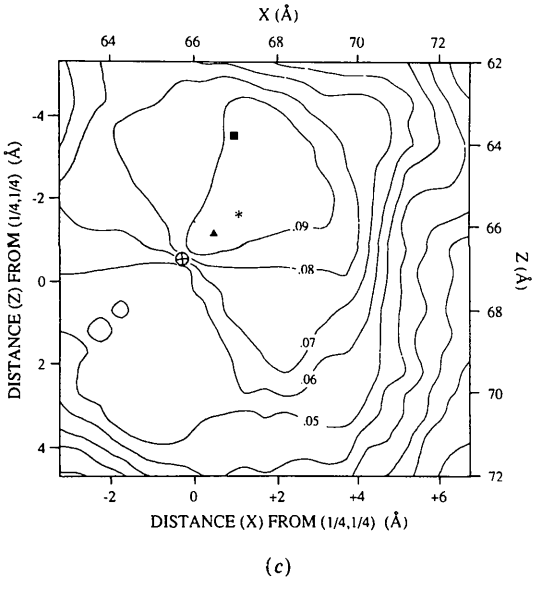
Using the particle position as refined at high resolution (Tsao, Chapman, Wu *et al.*, 1992), a search was made for the inner and outer radii (Fig. 7*a*) which gave the optimal agreement between model and 'true' phases for data from 30 to 13.5 Å resolution. Thus the data used here were the same as used in the original model parameter determination. Optimal was defined as that giving the highest absolute value of the |**F**|-weighted mean cosine phase difference, $\langle \cos(\Delta\varphi) \rangle$. The search indicated that the refined outer and inner radii that had been determined from the amplitudes (Fig. 3; Table 2) were within 1.0 and 3.5 Å,



(*a*)



(*b*)



(*c*)

Fig. 7. Retrospective searches for optimal models. |**F**|-weighted mean cosine phase differences between spherical-shell and back-extended phases are shown as functions of inner and outer radii in (*a*) and (*b*) and position in (*c*). The functions are contoured in steps of 0.01 with negative values shown in dashed lines. (*a*) and (*c*) were calculated for resolutions between 30 and 13.5 Å and (*b*) was calculated for 30 to 20 Å. In (*a*) and (*b*) are marked the points corresponding to the refined models and model 128, the model used in the actual structure determination of CPV (Tsao, Chapman, Wu *et al.*, 1992). In (*c*) are marked: the highest point (*); the position refined to 3.8 Å (▲) (Tsao, Chapman, Wu *et al.*, 1992); the position of $(\frac{1}{4}, 0, \frac{1}{4})$ (⊕); the position refined at 13.5 Å (■) with a spherical shell of outer radius 121.5 Å.

Table 4. *The quality of phases for spherical shell models*

Outer radius (Å)	Inner radius (Å)	Resolution of data (Å)	$\langle\cos\Delta\alpha\rangle$*	Comment
128.0	85.0	30–20	0.065	Used for CPV phase determination
121.5	81.7	30–20	0.127	Refined parameters for 30–13.5 Å data
122.0	85.0	30–20	0.152	Model with best agreement with true phases
121.5	81.7	30–13.5	0.075	Refined parameters for 30–13.5 Å data
122.0	85.0	30–13.5	0.099	Model with best agreement with true phases
130.0	91.7	30–13.5	−0.064	Refined parameters for 30–13.5 Å data (Babinet)
131.5	93.5	30–13.5	−0.076	Model with best agreement with the Babinet inverse of the true phases

* $\langle\cos\Delta\alpha\rangle$ is the $|F|$-weighted mean cosine phase difference between phases obtained from a spherical shell model and 'true' CPV phases.

respectively, of those that would have given the most accurate phases. When repeated within the resolution range 30 to 20 Å used for the initial CPV phase extension, spherical models can better approximate the 'true' phases (Table 4). Phases calculated from the model using parameters refined against the structure amplitudes were slightly worse than optimal but significantly better than the phases derived from the model used in the actual CPV phase extension.

The electron density of the nucleic acid

Unlike the radii and position, the refined nucleic acid density value can be validated only indirectly, because the 'true' phases were determined by back-extension in which the nucleic acid was set to zero for regions within 75 Å of the viral center. If the value of ρ'_{NA} is arbitrarily changed from 1.2 to 0.0, then the outer radius of the model that best fits the structure amplitudes increases by 4 Å (Fig. 8). The similarity of the refined radius (121.5Å) to that refined against crystallographic data for empty (DNA-lacking) varions (120.5 Å) (Wu, Keller, Agbandje, Tsao, Chapman & Rossmann, unpublished results) and to the radius (122 Å) that gives best agreement to the 'true' phases (calculated when ρ'_{NA} was set to zero) suggests that the refined value of ρ'_{NA} for full virions (Table 2) is at least approximately correct. This value of ρ'_{NA}, refined using data to 13.5 Å resolution (1.2), is lower than that when refined with data to 20 Å resolution (1.5) and also lower than that refined against solution scattering data (1.7). This last value agrees with that found from solution scattering data for cowpea mosaic virus (Schmidt *et al.*, 1983), while the decrease of ρ'_{NA} with increasing resolution is as expected if most of the nucleic acid is disordered.

Three models were tried for the initial CPV phasing trials (Tsao, Chapman, Wu *et al.*, 1992). Model 128* was consistent with packing considerations and the solution scattering. The parameters of models 122 and 132 were similar to those reported in this paper, but these models were selected at an earlier stage of refinement, when ρ'_{NA} was still fixed at 1.67. However, in all cases ρ'_{NA} was reset to zero prior to phase calculation and refinement. By comparing the agreement of structure-factor magnitudes (Fig. 8), it is clear that when ρ'_{NA} is set to zero, the optimal outer radius increases from 122 to 126 Å, making model 128 the best of the three. This suggests that, if ρ'_{NA} is to be set to zero for initial phase calculation, more appropriate radii would be selected through the imposition of this constraint during parameter refinement. However, as can be seen by comparing Fig. 8(*c*) with Fig. 7(*a*), the best model when $\rho'_{NA}=0$ would have a significantly poorer $\langle\cos(\Delta\varphi)\rangle$ (0.095) than optimal (0.152). This shows that it would be better to refine ρ'_{NA} and have independent nucleic acid and solvent densities during phase refinement.

When considering the effects of the nucleic acid region, examination of (9) shows that changing ρ'_{NA} from 1.16 to 0 introduces a mean change of 34% in the $\mathbf{F}_{calc}$. Note also that setting $R_i=0$ is the same as setting $\rho'_{NA}=1.0$ and corresponds to a mean change in $\mathbf{F}_{calc}$ of 4% for CPV. Thus, the refined inner radius of full virions will be much less precise than the outer radius (Figs. 7 and 8), but also less important in the initial phasing provided that the refined ρ'_{NA} value is used.

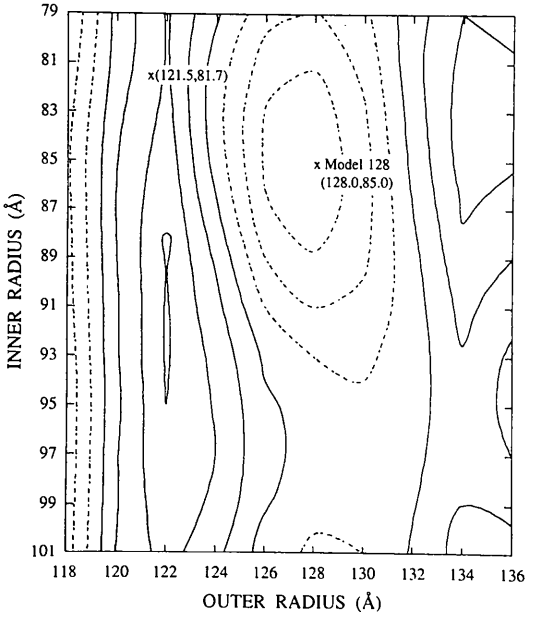
While retrospectively examining the success of model 128, it is interesting to examine why model 128 led to the Babinet opposite solution in spite of a positive correlation of phases between model 128 and back-extended phases for resolutions of 30 to 20 Å (Fig. 7*b*). A comparison of $\langle\cos\Delta\varphi\rangle$ as a function of resolution for models 122 and 128 shows that, in the 22 to 20 Å resolution range, the correlation becomes negative for model 128 (Fig. 9). Apparently, it is the narrow outermost shell that is most significant in the extension of phases to higher resolution. It is also of interest to note that model 128 proved successful, although, in retrospect, it is found that between 30 and 20 Å the initial phases had a mean random error of 85°. The success of the SIR extension, from an initial phasing of only 5% of the reflections (Tsao, Chapman, Wu *et al.*, 1992) attests to the power of the application of 60-fold noncrystallographic symmetry. It is not surprising, with such gross oversampling, that the iterative refinement methods used can correct a large random error. While large random errors are tolerable, the sensitivity of phase extension to model parameters (Tsao, Chapman & Rossmann, 1992) and

* The number indicates the outer radius, in Å, of the model.

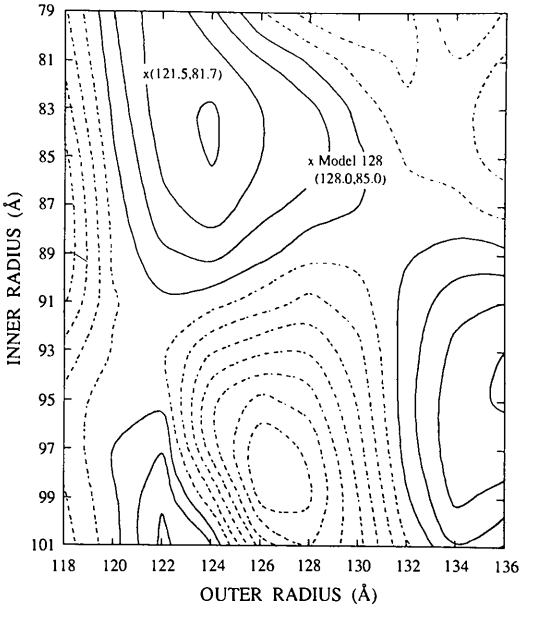
the difficulty of finding suitable models by trial and error (Tsao, 1990; McKenna *et al.*, 1991) indicate that it is important that the initial phase error is neither systematic nor consistent with an incorrect model.

Applicability to other structure determinations

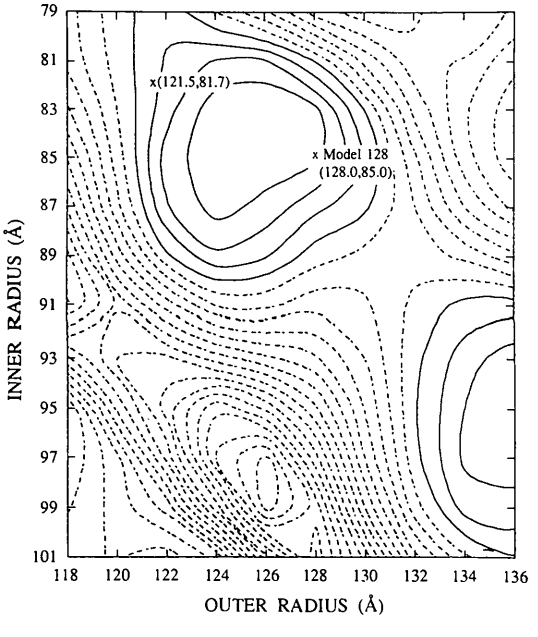
CPV had been thought to be an ideal candidate for this approach to phasing because negatively stained electron micrographs of CPV had revealed little surface feature, suggesting that it could be approximated by a spherical-shell model. In retrospect (Tsao *et al.*, 1991), CPV is seen to have considerable surface relief with spikes and depressions such that the distance from the center to the outer surface varies from 112 to 140 Å. If surface topology is more extreme, or if the viruses are oriented such that centricity cannot be removed by averaging, a more detailed (non-centric) model may be required, such as might be available from frozen hydrated electron microscopy images (reviewed by Chiu, 1986). Due to the sensitivity of the phases to the outer radius, it will be impor-



(*a*)



(*b*)



(*c*)

Fig. 8. Correlation coefficient (r) is shown as a function of model radii for (*a*) $\rho'_{NA} = 1.16$, (*b*) $\rho'_{NA} = 0.58$ and (*c*) $\rho'_{NA} = 0.0$.

$$r = \left[\sum((\langle|\mathbf{F}_{obs}|\rangle - |\mathbf{F}_{obs}|)(\langle|\mathbf{F}_s|\rangle - |\mathbf{F}_s|)\right] \times \left[\sum(\langle|\mathbf{F}_{obs}|\rangle - |\mathbf{F}_{obs}|)^2 \sum(\langle|\mathbf{F}_s|\rangle - |\mathbf{F}_s|)^2\right]^{-1/2}.$$

r is plotted in steps of 0.01 using solid lines above 0.23. These plots show that the best-fitting outer radius changes from 122 to 126 Å if the nucleic acid density is artificially set to zero.

tant to refine against the crystallographic data not only the position but also a magnification factor for the electron microscopy image.

It is possible that similar methods might be extended to other macromolecular assemblies, although viruses provide particularly amenable examples for several reasons. Refinement is facilitated because the virus can be represented by a model with few parameters. Furthermore, viruses have high (up to 60-fold) noncrystallographic redundancy enabling the refinement of very poor starting phases.

Tsao, Chapman, Wu *et al.* (1992) demonstrated that it is possible to solve a virus capsid structure at high resolution through the extension of phases calculated *ab initio* from a spherical-shell model. While it may be possible to find acceptable parameters by trial and error, at best, this is likely to be time-consuming. Here, it has been shown that good starting phases can be obtained by refinement of the model to the diffraction data. In the absence of data below 30 Å resolution, there may be several discretely different models whose fits to the data are of similar quality. An estimate of outer radius by some other biophysical method may enable distinction between these models, but all are likely to give initial phase sets that are within the convergence radius either of the true solution or its Babinet inverse. The collection and analysis of the data of isomorphous derivatives have often been rate limiting to the determination of viral capsid structures. The methods discussed in this paper could accelerate capsid structure determination by dispensing with the need for isomorphous derivatives.

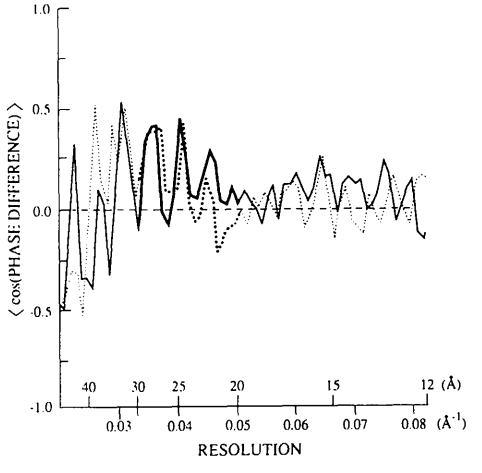


Fig. 9. The quality of phases calculated from model 122 (solid line) and model 128 (dotted line) with reference to back-extended phases, plotted as a function of resolution. The best phases are obtained within the resolution range of 20-30 Å. Within 40 Å resolution, where the data are poor and incomplete, the phases are poor. The correlation of model 128 and back-extended phases becomes negative at 22 Å resolution, explaining why phases extended from 20 Å resolution converged to the Babinet solution. The bold line shows that part of the resolution range from which phases were extended.

We are grateful to Mavis Agbandje and Kathy Smith for preparation of virus and crystals, Tim Baker for taking electron micrographs, those who helped in data collection at various synchrotrons (Jodi Bibler, Walter Keller, S. Krishnaswamy, Robert McKenna, Andrew Prongay, Tom Smith, Peter Willingmann, Di Xia and the staffs at CHESS, Daresbury, Brookhaven and DESY synchrotrons), Hao Wu for help in data processing and for calculation of the back-extended phases and Tim Schmidt for help in setting up the low-angle scattering experiment. The refinement program was adapted from a least-squares program assembled at UCLA in the laboratory of David Eisenberg. MSC gratefully acknowledges the help of Larry Weissman and Todd Yeates on whose subroutines parts of the original versions of the least-squares refinement were based. We are grateful to Lars Liljas, Robert McKenna and Sanjeev Munshi for stimulating discussions. We thank Helene Prongay and Sharon Wilder for help in the preparation of the manuscript. This work was supported by grants from the National Institutes of Health, National Science Foundation and the Lucille P. Markey Charitable Trust.

References

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

CHAPMAN, M. S., GIRANDA, V. L. & ROSSMANN, M. G. (1991). *Sem. Virol.* **1**, 413-427.

CHIU, W. (1986). *Annu. Rev. Biophys. Biophys. Chem.* **15**, 237-257.

DIAMOND, R. (1971). *Acta Cryst.* A**27**, 436-452.

FINCH, J. T. & HOLMES, K. C. (1967). *Methods Virol.* **3**, 351-474.

HARRISON, S. C. (1969). *J. Mol. Biol.* **42**, 457-483.

HARRISON, S. C. (1990). In *Virology*, edited by B. N. FIELDS & D. M. KNIPE, Vol. 1, pp. 37-61. New York: Raven Press.

HARRISON, S. C., OLSON, A. J., SCHUTT, C. E., WINKLER, F. K. & BRICOGNE, G. (1978). *Nature (London)*, **276**, 368-373.

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

HOSUR, M. V., SCHMIDT, T., TUCKER, R. C., JOHNSON, J. E., GALLAGHER, T. M., SELLING, B. H. & RUECKERT, R. R. (1987). *Proteins*, **2**, 167-176.

JACK, A. & HARRISON, S. C. (1975). *J. Mol. Biol.* **99**, 15-25.

JOHNSON, J. E., AKIMOTO, T., SUCK, D., RAYMENT, I. & ROSSMANN, M. G. (1976). *Virology*, **75**, 394-400.

KIM, S., SMITH, T. J., CHAPMAN, M. S., ROSSMANN, M. G., PEVEAR, D. C., DUTKO, F. J., FELOCK, P. J., DIANA, G. D. & MCKINLAY, M. A. (1989). *J. Mol. Biol.* **210**, 91-111.

LILJAS, L. (1986). *Prog. Biophys. Mol. Biol.* **48**, 1-36.

LILJAS, L., UNGE, T., JONES, T. A., FRIDBORG, K., LÖVGREN, S., SKOGLUND, U. & STRANDBERG, B. (1982). *J. Mol. Biol.* **159**, 93-108.

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

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., OLSEN, N. H., BAKER, T. S. & INCARDONA, A. L. (1991). *Nature (London)*. In the press.

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

PATTISON, J. R. (1990). In *Virology*, edited by B. N. FIELDS & D. M. KNIPE, Vol. 2, pp. 1765-1784. New York: Raven Press.

RAYMENT, I., BAKER, T. S. & CASPAR, D. L. D. (1983). *Acta Cryst.* B39, 505-516.
RAYMENT, I., BAKER, T. S., CASPAR, D. L. D. & MURAKAMI, W. T. (1982). *Nature (London)*, **295**, 110-115.
ROSSMANN, M. G. (1990). *Acta Cryst.* A46, 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. & JOHNSON, J. E. (1989). *Annu. Rev. Biochem.* **58**, 533-573.
SCHMIDT, T., JOHNSON, J. E. & PHILLIPS, W. E. (1983). *Virology*, **127**, 65-73.
SMITH, M. H. (1968). In *Handbook of Biochemistry, Selected Data for Molecular Biology*, edited by H. A. SOBER, pp. C-28-C-35. Cleveland: CRC Press.
TSAO, J. (1990). PhD thesis, Purdue Univ., West Lafayette, Indiana, USA.
TSAO, J., CHAPMAN, M. S., AGBANDJE, M., KELLER, W., SMITH, K., WU, H., LUO, M., SMITH, T. J., ROSSMANN, M. G., COMPANS, R W. & PARRISH, C. R. (1991). *Science* **251**, 1456-1464.
TSAO, J., CHAPMAN, M. S. & ROSSMANN, M. G. (1992). *Acta Cryst.* A48, 293-301.
TSAO, J., CHAPMAN, M. S., WU, H., AGBANDJE, M., KELLER, W. & ROSSMANN, M. G. (1992). *Acta Cryst.* B48, 75-88.
VALEGÅRD, K., LILJAS, L., FRIDBORG, K. & UNGE, T. (1990). *Nature (London)*, **345**, 36-41.
WOBBE, C. R., MITRA, S. & RAMAKRISHNAN, V. (1984). *Biochemistry*, **23**, 6565-6569.

Acta Cryst. (1992). A48, 312-322

Some Applications of the Phased Translation Function in Macromolecular Structure Determination

BY G. A. BENTLEY AND A. HOUDUSSE

Unité d'Immunologie Structurale (CNRS URA 359), Institut Pasteur, 25 rue du Dr Roux, Paris 75724, France

(*Received* 14 *February* 1991; *accepted* 5 *November* 1991)

Abstract

Although the phased translation function was first described some time ago [Colman, Fehlhammer & Bartels (1976). In *Crystallographic Computing Techniques*, edited by F. R. Ahmed, K. Huml & B. Sedláček, pp. 248-258. Copenhagen: Munksgaard], it has been little used, especially in the application of molecular replacement to macromolocular structures. Nevertheless, the procedure is relatively easy to apply and deserves wider use. In this paper the versatility of the phased translation function in a number of different applications is examined and experience gained in obtaining optimal results in protein structure determination by this method is reported. Examples given show how it can be used to position an oriented fragment, to locate independent components with respect to a common crystallographic origin and to choose correctly between enantiomorphic space groups. Its performance is compared with other translation functions in common use.

1. Introduction

Molecular replacement is widely used to determine macromolecular structures since there now exists a large resource of known structures which may contain one or more examples closely homologous to the molecule under study. The method proceeds in two stages. Firstly, the homologous molecule of known structure, or a fragment of it, is oriented in the unit cell of the unknown structure by means of a rotation function. Secondly, the correctly oriented homologue is positioned with respect to a crystallographic origin using a translation function, providing an initial model of the unknown crystal structure for further refinement. Whereas the rotation functions in current use exploit properties of the Patterson function (Rossmann & Blow, 1962; Huber, 1965; Crowther, 1972), translation functions may use the Patterson function (Crowther & Blow, 1967; Harada, Lifchitz, Berthou & Jolles, 1981), the correlation between the observed and calculated structure amplitudes, as in the *R*-factor search (Taylor & Morley, 1959; Fujinaga & Read, 1987), or phased structure amplitudes, as in the phased translation function (Colman, Fehlhammer & Bartels, 1976; Doesburg & Beurskens, 1983; Read & Schierbeek, 1988; Cygler & Desrochers, 1989).

We have used the phased translation function (PTF) to aid the solution of two crystal structures of Fab fragments by molecular replacement: the complex FabD1.3-FabE225 (Bentley, Boulot, Riottot & Poljak, 1990) and FabE5.2 (Houdusse, Bentley, Boulot, Eiselé & Poljak, unpublished results). Although we used it primarily to locate independent components with respect to a common origin, we performed additional calculations to test its general

0108-7673/92/030312-11$03.00 © 1992 International Union of Crystallography