Acta Cryst. (1980). A**36**, 747–754

Procedures for Detection and Idealization of Non-crystallographic Symmetry with Application to Phase Refinement of the Satellite Tobacco Necrosis Virus Structure

BY C. E. NORDMAN

Department of Chemistry, University of Michigan, Ann Arbor, Michigan 48109, *USA*

(*Received* 2 *January* 1980; *accepted* 11 *March* 1980)

Abstract

An electron-density function with a region of local non-crystallographic symmetry is sampled with a movable symmetry grid. This provides a means of optimizing the parameters of the locally symmetric region as well as of idealizing its electron density through symmetry averaging. An iterative procedure for phase refinement uses local symmetry averaging and Fourier inversion, with selective reciprocal-space damping of unmodified function regions, and correction for solvent scattering. Application to the crystal-structure analysis of STNV, with 60-fold icosahedral non-crystallographic symmetry, shows good convergence and capability of phase extension from 10 to 4 Å resolution.

Introduction

Methods for determining the nature, directions and point of intersection of the local symmetry axes in regions of non-crystallographic symmetry are well established (Rossmann, 1972). This knowledge has been used to improve the interpretability of electron-density maps and as an aid in locating heavy atoms.

However, the most powerful direct-space use of such local symmetry information is the generation of phases by Fourier inversion of locally symmetry-averaged electron-density maps, and the iterative phase refinement of such maps. Programs for accomplishing this have been developed, notably by Bricogne (1976), who has also discussed the optimization and limitations of the process, as well as earlier work in this field. Independently, Johnson (1978) has developed a similar procedure of iterative direct-space symmetry averaging and Fourier inversion.

These methods have played a central part in the solution and refinement of the three largest crystallographic problems to reach interpretable resolution, namely the protein disk of Tobacco Mosaic Virus (Bloomer, Champness, Bricogne, Staden & Klug, 1978), Tomato Bushy Stunt Virus (Harrison, Olson, Schütt, Winkler & Bricogne, 1978) and Southern Bean Mosaic Virus (Suck, Rayment, Johnson & Rossmann, 1978; Abad-Zapatero, Abdel-Meguid, Johnson, Leslie, Rayment, Rossmann, Suck & Tsukihara, 1980).

In these applications the phase information derived from non-crystallographic-symmetry averaging has been used in combination with multiple-isomorphous-replacement data, or to improve the phases of maps initially phased by multiple-isomorphous-replacement methods.

Argos, Ford & Rossmann (1975), working with the known structure of glyceraldehyde-3-phosphate dehydrogenase, with non-crystallographic molecular symmetry 222, demonstrated the potential value of the method in phase extension, *i.e.* the determination of *ab initio* phases in a resolution range beyond that for which phase information is initially available. Johnson, Akimoto, Suck, Rayment & Rossmann (1976) applied the method to Southern Bean Mosaic Virus, extending initial 35 Å phases based on a spherical-particle model to 22·5 Å by icosahedral averaging.

Satellite Tobacco Necrosis Virus (STNV) is a small icosahedral virus of molecular weight 1·7 million. It crystallizes in the monoclinic space group $C2$ with four molecules per cell. The crystallographic asymmetric unit is an entire virus particle, thus the full 60-fold symmetry of the icosahedral point group 532 is non-crystallographic (Strandberg, Vaara, Unge, Fridborg, Kannan, Borell, Lentz & Nordman, 1978; Strandberg, Unge, Liljas, Vaara, Kannan, Fridborg, Borell & Nordman, 1979; Unge, Liljas, Strandberg, Vaara, Kannan, Fridborg, Nordman & Lentz, 1980).

This paper describes a set of computational procedures for utilizing non-crystallographic symmetry, developed in the course of the structure analysis of STNV. The programs are derived from Patterson search routines previously in use in this laboratory (Nordman, 1966; Schilling, 1970), and employ function-handling subroutines directly adapted from the latter. The programs also have much in common with those of Bricogne (1976) and Johnson (1978).

Another purpose of this paper is to account for experiences gained in the use of these methods in phase extension with native-only STNV data.

0567-7394/80/050747-08$01.00

Function storage and retrieval

The storage of the electron-density function in the computer memory, and the retrieval of interpolated values are done in a manner identical to that used in Patterson search applications (Schilling, 1970). The function is sampled on a crystallographic, *i.e.* fractional unit-cell coordinate, grid, covering one asymmetric unit. The function values are integers in the range 0 to $2^n - 1$, where n, the number of bits per value, may be chosen according to the accuracy desired and the available core storage. In the work described here $n = 8$ was used. This choice is particularly convenient on IBM-type machines, since it allows the function to be stored as a three-dimensional array of individually addressable logical variables. On other machines several values would be packed into one word. In the following we refer to the array of stored function values as the *map*. The map covers one crystallographic asymmetric unit.

A function value at an arbitrary point in the unit cell is retrieved by interpolation in the map. A space-group-dependent section in the map-handling subroutine first moves the point into the asymmetric unit covered by the map. The desired function value is then evaluated by one of several interpolation schemes available in the subroutine.

For most purposes for which this function-storage and -retrieval system has been used, an eight-point interpolation scheme has been employed. Here the eight map values at the corners of the 'box' surrounding the function point are used in a manner so as to give a continuous representation of the interpolated function. Let $\rho_{000}, \rho_{100}, \rho_{010}, \ldots, \rho_{111}$ be the tabulated map values at the eight corners of the box, and let ρ_{000} be the corner *nearest* the point at which the function is to be evaluated. If u, v and w are the local coordinates of this point expressed in fractions of the box edges and referred to the ρ_{000} corner as origin, it follows that $0 \leq u(v, w) \leq 0{\cdot}5$, *i.e.* $u(v, w)$ are taken in the positive or negative direction of the crystal coordinates $x(y, z)$ depending on the octant of the box in which the point lies. The interpolated function value is expressed as an eight-term polynomial in u, v and w, where the coefficients of the respective powers are:

$$
\begin{aligned}
1 \quad & \rho_{000}; \\
u \quad & -\rho_{000} + \rho_{100}; \\
v \quad & -\rho_{000} + \rho_{010}; \\
w \quad & -\rho_{000} + \rho_{001}; \\
uv \quad & \rho_{000} - \rho_{100} - \rho_{010} + \rho_{110}; \\
vw \quad & \rho_{000} - \rho_{010} - \rho_{001} + \rho_{011}; \\
uw \quad & \rho_{000} - \rho_{100} - \rho_{001} + \rho_{101}; \\
uvw \quad & -\rho_{000} + \rho_{100} + \rho_{010} + \rho_{001} - \rho_{110} - \rho_{011} - \rho_{101} \\
& +\rho_{111}.
\end{aligned}
$$

More rapidly computable approximations, not used in the STNV work, employ only the first four, or the first one, of these eight terms.

Symmetry search and optimization

The routine described in the preceding section lends itself to several applications. One is the calculation of arbitrary sections, plane or otherwise, through an electron-density function.

Another is the previously mentioned search of Patterson functions for sets of interatomic vectors corresponding to some structural fragment known to be present in the structure (Nordman, 1966; Schilling, 1970). Similarly, searches of poorly phased or incompletely resolved electron-density maps for sets of atomic position vectors have been found useful. The criterion of fit in all these searches is some function of the vector or electron densities at the vector points, expressing the degree to which all search vectors fall on satisfactorily high ground in the Patterson or electron-density function.

In the presence of local *non-crystallographic symmetry*, a closely related calculation can be used to find the directions of local symmetry axes or planes and, where applicable, the point of intersection of such axes or planes. In this calculation the sampling of the stored function is done with a *symmetry grid* instead of the above-mentioned search vectors.

The symmetry grid is a set of points arranged so as to display the exact symmetry of the point group in question. Let the order of the local non-crystallographic point group be N_s. The symmetry grid then consists of N_s (non-crystallographic) asymmetric units, which we shall call *subunits*. Each subunit is delimited by a set of planes, through the point-group origin, which mark the inter-subunit boundaries. An outer and, if applicable, an inner boundary describe the outer (inner) limits of the non-crystallographically symmetric region. The symmetry grid is similar to the intermediate Cartesian grid described by Bricogne (1976) as an alternative method of symmetry averaging.

In the case of the icosahedral point group 532 N_s equals 60, and the subunit of the symmetry grid may be taken as the triangular wedge having as its edges one threefold axis and two adjacent fivefold axes. The relationship of this grid subunit to the shape of the chemical subunit in the viral coat is immaterial.

The points of the symmetry grid are laid out so as to fill one grid subunit in an approximately uniform manner. The grid-generating subroutine accepts a specifiable step parameter expressing the maximum distance between adjacent grid points. Subject to this restraint the subunit is filled with the minimum number of grid points, N_i. The value of N_i is returned to the calling program, and the Cartesian coordinates of the N_i points are stored. The subunit in question is taken as subunit 1.

A region of the function presumed to have local non-crystallographic symmetry is sampled by rotating and translating the symmetry grid into the region. Function values at the points of the symmetry grid are

evaluated by interpolation in the map. Let $\rho(\mathbf{x}_{is})$ be the function value at the location of the ith point of symmetry-grid subunit s.

We now evaluate

$$\sigma_i^2 = (1/N_s) \sum_{s=1}^{N_s} [\rho(\mathbf{x}_{is}) - \bar{\rho}_i]^2, \quad (1)$$

where

$$\bar{\rho}_i = (1/N_s) \sum_{s=1}^{N_s} \rho(\mathbf{x}_{is}). \quad (2)$$

A low value of σ_i^2 indicates a high degree of symmetry at the symmetry grid points $\mathbf{x}_{i1}, \mathbf{x}_{i2}, \ldots, \mathbf{x}_{iN_s}$. The sum

$$\sigma^2 = \sum_{i=1}^{N_i} \sigma_i^2 \quad (3)$$

is a measure of the presence of symmetry of the point group in question in the entire region sampled by the symmetry grid. A low (or zero) value of σ^2 represents high (or perfect) symmetry.

Only the coordinates of the N_i points in subunit 1 of the symmetry grid need be explicitly stored. Having rotated these points through the Eulerian angles representing the orientation of the particle in the cell, the coordinates of each are multiplied, in turn, by N_s matrices of the type $\mathbf{CR}(s)$, where $\mathbf{R}(s)$ first rotates subunit 1 into subunit s, and $\mathbf{C}$ transforms the Cartesian coordinates to fractional crystal coordinates. Thus the transformation of the Cartesian coordinates of grid point i in grid subunit 1 into fractional coordinates of point i in subunit s is accomplished with one matrix multiplication.

To refine the coordinates of the center of the symmetric region, or the spatial orientation of its symmetry elements, small adjustments are made in these parameters so as to minimize σ^2 of (3). The procedure may be employed for purposes similar to the rotation function (Rossmann & Blow, 1962) by placing the symmetry-grid center at the origin of the Patterson function, and performing a systematic, in general three-dimensional, search of Euler-angle space over a region determined by the (angular) size of the subunit in the point group in question.

A third application of this sampling procedure involves a display of σ_i^2 (equation 1) as a function of $\mathbf{x}_i$, the position of the point i in the subunit. This procedure can be helpful in identifying the boundary of a non-crystallographically symmetric region. Points outside the symmetric region have σ_i^2 values which are no lower than those of N_s randomly chosen points in the function.

Symmetry averaging

By 'symmetry averaging' we mean the idealization of a local approximate (non-crystallographic) symmetry present in a given region of the unit cell.

The rotatable and translatable symmetry grid, described in the preceding section, lends itself to a simple and computationally economical procedure for symmetry averaging. We assume that the symmetry grid has been positioned so as to minimize the sum (3). In the symmetry averaging process the map values are modified in such a way as to make (3) vanish, or nearly so.

The computation consists of three steps, each carried out in the order of the points of the symmetry grid, with the index s changing fastest.

In the first step the coordinates of the map point $\mathbf{m}_{is}^0$ nearest the current grid point $\mathbf{x}_{is}$ are stored, as are flags identifying the three second-nearest map points $\mathbf{m}_{is}^x$, $\mathbf{m}_{is}^y$ and $\mathbf{m}_{is}^z$. The coordinates of these second-nearest points differ from the 'nearest' map point by ± 1 map step. The value of $\rho(\mathbf{x}_{is})$, the value of

$$\Delta\rho_{is}^0 = \rho(\mathbf{m}_{is}^0) - \rho(\mathbf{x}_{is})$$

and the three second-nearest values

$$\Delta\rho_{is}^x = \rho(\mathbf{m}_{is}^x) - \rho(\mathbf{x}_{is}),$$

and analogously $\Delta\rho_{is}^y$ and $\Delta\rho_{is}^z$, are also stored. When s equals N_s, $\bar{\rho}_i$ (equation 2) is evaluated and written on an external file along with the N_s sets of $\mathbf{m}^0$ coordinates, second-nearest map-point flags and $\Delta\rho_{is}$ values. The values of $\rho(\mathbf{x}_{is})$ are not needed. This operation is repeated for all values of i.

In the second step the external file is read. For each i, the N_s quantities $\bar{\rho}_i + \Delta\rho_{is}^0$ are inserted at their respective 'nearest' map points $\mathbf{m}_{is}^0$, replacing the previous values $\rho(\mathbf{m}_{is}^0)$ at these points. When the substitution is made at map point $\mathbf{m}_{is}^0$, a one-bit flag is set to 1 at the location corresponding to $\mathbf{m}_{is}^0$ in an initially zeroed bit array having one bit for each element in the map array.

It should be noted that there is no guarantee that all map points within the region covered by the symmetry grid will have been substituted at the end of this step. This would be achieved only if the symmetry grid step were chosen sufficiently small in relation to the interval between adjacent map points. In that case many map points would be substituted twice, or more, since the likelihood that a given map point would be the nearest one to several grid points increases with the density of the grid. Such multiple substitution would be wasteful, but otherwise harmless.

An efficient compromise involves substitution of the second-nearest map-point data. This is done in the third step. The external file is read again. This time the $3N_s$ quantities $\bar{\rho}_i + \Delta\rho_{is}^x$ (and similarly for y and z) are used. However, substitution at map points $\mathbf{m}_{is}^x$ (and y, z) is made only if the bit-array entry indicates that the map point has not already been substituted, in step 2. Thus the 'second-nearest' data are invoked only at those map points which would otherwise have been missed.

The entire calculation is proportional to $N_s N_i$. Since N_s is fixed by the point-group symmetry, the disposable parameter is N_i, the number of points in the subunit of the symmetry grid, inversely proportional to the third power of the grid step. The proportion of first- and second-nearest substitutions depends on the ratio of this grid step to the map point interval. As an example, in the symmetry averaging of the STNV map at 4 Å resolution it was found that a grid step of 1·05 Å and a map point interval of 1·30 Å resulted in $0{\cdot}978 \times 10^6$ nearest and $0{\cdot}075 \times 10^6$ second-nearest substitutions.

In the STNV symmetry averaging at 4·0 Å resolution, it was found advantageous to include three additional quadratic terms in the polynomial for the interpolated function. If the function value given by the eight-point interpolation scheme is denoted ρ^8, then

$$\rho^{11} = \rho^8 + k\Delta\rho,$$

where

$$\begin{aligned}\Delta\rho = {} & (\rho_{000} - 0{\cdot}5\rho_{100} - 0{\cdot}5\rho_{\bar{1}00})\, u(1-u) \\ & + (\rho_{000} - 0{\cdot}5\rho_{010} - 0{\cdot}5\rho_{0\bar{1}0})\, v(1-v) \\ & + (\rho_{000} - 0{\cdot}5\rho_{001} - 0{\cdot}5\rho_{00\bar{1}})\, w(1-w)\end{aligned}$$

represents an eleven-point interpolated value based on the eight corners of the box, and the three points $\rho_{\bar{1}00}$, $\rho_{0\bar{1}0}$ and $\rho_{00\bar{1}}$ outside the box. The quantity k was treated as an adjustable parameter in the range $0 \leq k \leq 1$. Its value was optimized by symmetry averaging a given map, and then evaluating the resulting symmetry (equation 3) using a symmetry grid rationally unrelated to the first. It was found that $k = 0{\cdot}4$ produced the lowest σ^2 (equation 3), and that this value was significantly lower than that produced by the eight-point interpolation, for which $k = 0$.

Phase refinement and solvent correction with application to STNV

In the structure analysis of STNV, a phase-refinement procedure was used, based on symmetry averaging of an F_o electron-density synthesis followed by calculation of improved phases, by Fourier inversion, for input into the F_o map of the next cycle.

Given an F_o map, the first step is to replace the function values at the 'background' points in the map array by their mean value. A background point is one which does not fall inside the outer envelope of any particle (or molecule) in the structure. A separate program generates a file containing the coordinates of all background points in the crystallographic asymmetric unit. The contents of this file depend on the center coordinates of the particle, and on the geometry of the outer envelope of the particle. The background file is updated each time the particle position or orientation is changed.

The background boundary, or particle envelope, and the outer boundary of the symmetrization region are not necessarily identical. While the background boundary is chosen to reflect the best estimate of the particle envelope, the symmetrization boundary must be such as to ensure that no symmetrization regions of neighboring particles overlap.

The next step in the phase-refinement cycle is the symmetry averaging, described in the preceding section. In the case of a virus, the central nucleic acid cannot be expected to exhibit the symmetry of the protein coat. Accordingly, an inner as well as an outer symmetrization boundary is used. At low resolution the assignment of the inner boundary is necessarily somewhat arbitrary. An approximate assignment of this boundary can be made by calculating the electron-density integral, as a function of radial distance from the particle center, and apportioning the nucleic acid and protein regions to correspond to the chemically known electron content of each region. In the case of STNV an apparent, non-spherical nucleic acid–protein interface began to emerge at about 6 Å resolution. The inner symmetrization boundary was successively revised as this interface became more clearly traceable.

When the symmetry averaging has been completed, any map points within the particle envelopes, which have a value less than the averaged background value, are set to that value.

In the next step the background-averaged and symmetrized map is Fourier inverted (Ten Eyck, 1973) to yield a set of 'calculated' structure factors, F_c, to a specifiable upper limit in $\sin\theta/\lambda$. R values and other agreement indices between the $|F_c|$ values and the measured $|F_o|$ values are computed as functions of $\sin\theta/\lambda$. The F_o Fourier synthesis for the next phase-refinement cycle is computed with $|F_o|$ magnitudes and F_c phases. Accidentally unrecorded reflections, *i.e.* reflections within the resolution limit whose $|F_o|$ values are unknown, are assigned the structure amplitude $|F_c|$.

The preceding outline is a somewhat simplified description of the phase-refinement cycle. Two additional steps in the computation are a solvent scattering correction and a partial structure-factor damping feature. These steps are described in the following paragraphs.

It is a common experience in macromolecular crystallography that $\overline{|F_c|}/\overline{|F_o|}$ for low-angle reflections tends to exceed unity, unless a correction is made for the contribution of the solvent to the Bragg intensities (Moews & Kretsinger, 1975). This behavior was clearly displayed by the uncorrected STNV data, with $\overline{|F_c|}/\overline{|F_o|} \simeq 3$–4 for the lowest-angle reflections. Another symptom of solvent scattering was a tendency for a large fraction, about 25%, of the map points

within the particle envelope to have function values less than the background average, a physically unreasonable condition.

The following procedure was used to correct approximately for the solvent scattering. A conversion factor relating the true electron density, in e $Å^{-3}$, to the map values above the background was calculated as the ratio of the electron content of the virus particle, taken as 9×10^5 electrons, to the volume integral of the map function above background. With this conversion factor the map value corresponding to the true electron density of water (0·335 e $Å^{-3}$) was calculated. An artificial 'electron density' map was constructed with the map equivalent of 0·335 e $Å^{-3}$ inside the particle envelopes, and the background average value elsewhere. The Fourier inversion of this map yields, by Babinet's principle, an approximate set of $-F_w(\mathbf{h})$ values, where $F_w(\mathbf{h})$ is the contribution of the inter-particle water to the Bragg reflection **h**. An essentially equivalent procedure, applicable at low resolution and not requiring accurate knowledge of the particle envelope, is to truncate the background-averaged and symmetrized electron-density function at the map equivalent of 0·335 e $Å^{-3}$, and invert this function to obtain the $-F_w(\mathbf{h})$.

If the $F_c(\mathbf{h})$ is the structure factor calculated by inversion of the intact, symmetrized electron-density function, then $F_c(\mathbf{h}) + F_w(\mathbf{h})$ approximates the resultant of the particle and inter-particle solvent scattering. A temperature factor applied to the solvent component according to

$$F_{c,\mathrm{wet}}(\mathbf{h}) = F_c(\mathbf{h}) + F_w(\mathbf{h}) \exp[-B_w(\sin\theta/\lambda)^2] \quad (4)$$

has been found advantageous. The resulting solvent-corrected calculated structure factors show sharply improved agreement with the experimental $|F_o|$ values at low angles. The correction falls to insignificant values at about 7 Å resolution. Here $B_w = 500$ $Å^2$ was chosen by trial to optimize R.

The comparison between the $|F_o|$ values and the corrected, 'wet' $|F_c|$ values (equation 4) is done without rescaling of either set. The resulting values of

$$\Delta F(\mathbf{h}) = |F_o(\mathbf{h})| - |F_{c,\mathrm{wet}}(\mathbf{h})|$$

are added to the 'dry' $|F_c(\mathbf{h})|$ values to give a set of 'dry' $|F_o|$ values. These magnitudes with the $F_c(\mathbf{h})$ phases are used as input to the next phase-refinement cycle.

Parallel refinement tests of STNV were run at 8 Å resolution with and without the solvent correction. The results showed the essentially complete elimination of below-background regions within the particle envelope when the solvent correction was applied. Improved adherence to symmetry in the symmetrization region and constancy in the background region were also noted. For these reasons the solvent-corrected electron-density function was taken as the best available representation of the non-solvent structure for the purpose of symmetry averaging.

In the phase-refinement cycle outlined above, three distinct treatments of map regions are applied. The background region is averaged, the symmetrization region is symmetry-averaged, and the remaining regions, if any, are left unmodified. The latter regions, if present, are the nucleic-acid region and, possibly, a thin layer between the particle envelope and the outer boundary of the symmetrization region.

It is clear that the convergence of the phase refinement must be endangered if the unmodified region is large in relation to the modified regions, depending perhaps also on the symmetry number N_s applicable in the symmetrization region. In trial runs with poor starting phases serious non-convergence was indeed occasionally experienced in the form of a build-up of excessive 'structural detail' in the unmodified regions.

To insure against this behavior an additional, optionally invoked, Fourier inversion step was added to the phase refinement cycle. Following the Fourier inversion of the background-averaged and symmetry-averaged map, to yield $F_c(\mathbf{h})$ values, the map values in the symmetrized region are set to the background value. Thus only the unmodified parts of the map have values greater than the background value. Fourier inversion of this map yields a set of structure-factor contributions $F_n(\mathbf{h})$, the 'nucleic acid' components of the $F_c(\mathbf{h})$. By Wilson statistics, it is now ascertained that the fall-off of $|F_n(\mathbf{h})|$ with $\sin^2\theta/\lambda^2$ is at least as steep as the fall-off of $|F_c(\mathbf{h})|$. If this is not the case, an appropriate damping factor $T = \exp(-B_n \sin^2\theta/\lambda^2)$ is applied to the 'nucleic acid' component of $F_c(\mathbf{h})$ by adding the quantity $(T-1)F_n(\mathbf{h})$ to each $F_c(\mathbf{h})$.

In the normal course of phase refinement at 4 Å resolution, the damping feature is not generally invoked. That is, the natural thermal fall-off of the nucleic-acid component is slightly steeper than that of the entire particle indicating, perhaps, some structural disorder in the nucleic-acid region.

The behavior of the phase refinement in its initial stages is illustrated in Fig. 1. Electron-density sections are shown perpendicular to a particle three-fold axis. The leftmost map is the double-isomorphous-replacement map at 10 Å resolution computed with about 500 low-angle reflections missing out of the total of 8600. The next map is the result of one cycle of background averaging, symmetry averaging and Fourier inversion. The third and fourth maps are the results of four and twelve cycles of phase refinement, respectively.

A test of the convergence of this phase-refinement procedure at 10 Å resolution is shown in Figs. 2 and 3. The map obtained after several cycles of refinement of the original DIR map is shown in Fig. 2(*c*). In order to obtain starting phases for the test, this map was spherically symmetry averaged over the entire particle and the resulting spherically symmetrical map was

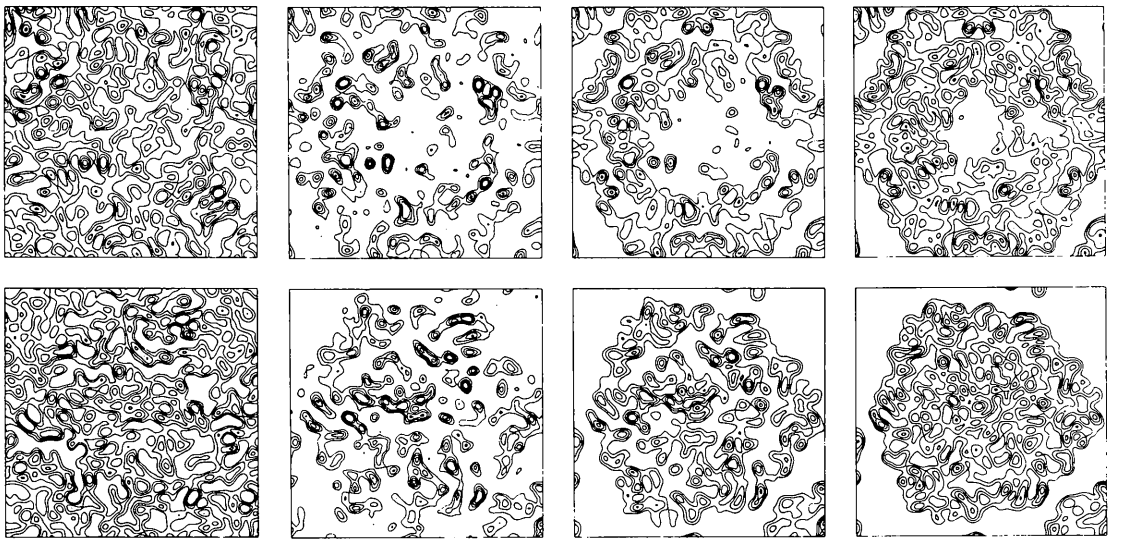

Fig. 1. Electron-density sections, at 10 Å resolution, perpendicular to a particle three-fold axis, through the particle center (top), and 40 Å from the center (bottom). Pairs from left to right show the unaveraged DIR map and maps calculated after one, four and twelve cycles of phase refinement.

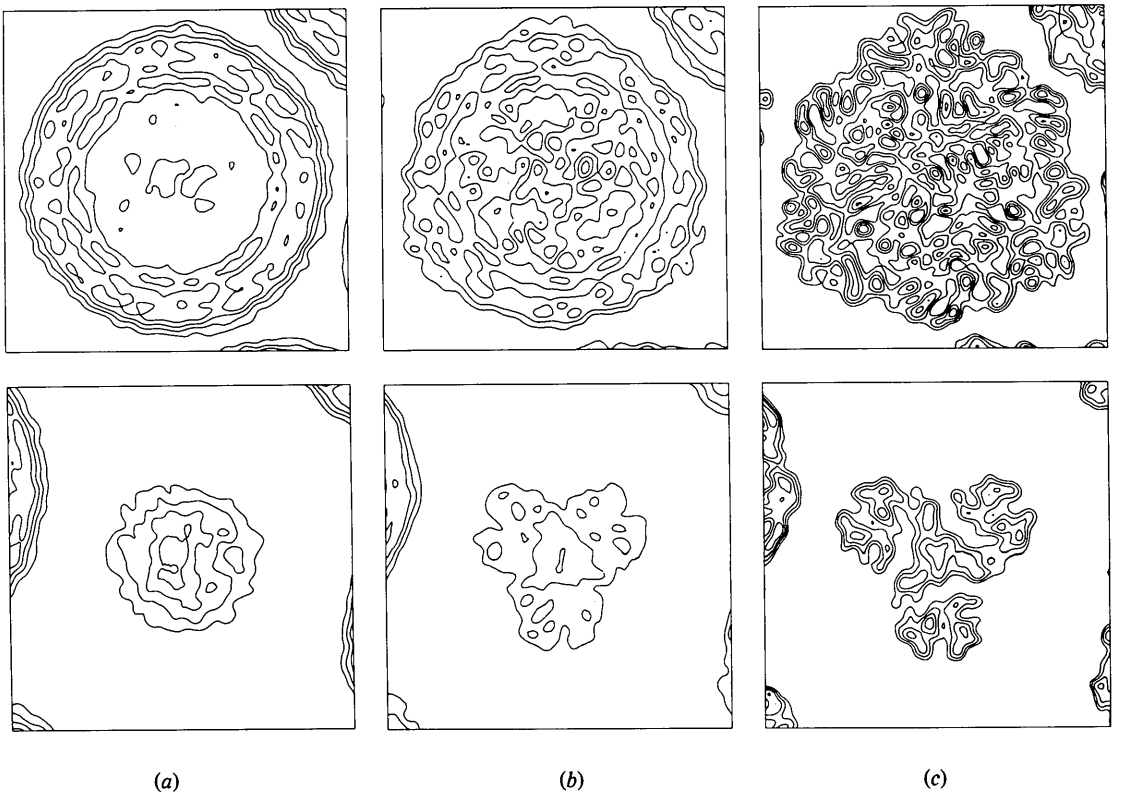

Fig. 2. Refinement of spherically averaged electron density at 10 Å resolution. Maps are perpendicular to a particle three-fold axis, at 40 (top) and 80 Å (bottom) from the particle center, following (*a*) one and (*b*) seven cycles of phase refinement. The 'best' refined map at 10 Å, based on DIR phases and 12 cycles of phase refinement is shown in (*c*).

inverted. Figs. 2(*a*) and (*b*) show, respectively, the maps after one and seven cycles of phase refinement. Fig. 3 shows the phase differences between the refined (Fig. 2*c*) model on one hand and the spherical, spherical +7 cycles and spherical +8 cycles models on the other. From Figs. 2(*a*)–(*c*) and 3 it is clear that a phase refinement towards the refined model is taking place.

Phase-extension experience with STNV

Beginning with the 10 Å set of phases from double isomorphous replacement, two series of phase refinement with stepwise increasing resolution were carried out. Each series used only native data; no phase information from heavy-atom derivatives beyond 10 Å resolution was used.

The first series used precession-camera data to a nominal resolution of 5·5 Å. This data set contained 44 600 measured reflections out of a total of 57 000 to the 5·5 Å resolution limit. The missing data were mainly in deep 'pie slices' of unrecorded data resulting from incomplete coverage of reciprocal space by the precession geometry.

Following phase refinement at 10 Å resolution (Fig. 1), data were successively added to the set, increasing the resolution in steps of 0·5 Å.

The addition of each shell of new data was initiated at the Fourier-inversion stage, by generating $F_c(\mathbf{h})$ values for the data to be added. These 'new' F_c values, invariably of low magnitude, were treated as 'unrecorded' reflections in the subsequent $F_o(\mathbf{h})$ Fourier synthesis. That is, they were entered as F_c values, not as F_o values, to avoid imposing the full $|F_o|$ magnitude on the presumably inaccurate *ab initio* phases. In the next cycle the $|F_c|$ values of these reflections, now generally higher than in the preceding cycle, were replaced with $|F_o|$, provided of course that a measured value had been recorded for the reflection in question. At the end of the series several cycles of phase refinement were done at 5·5 Å resolution, giving an R of 0·20 for the measured reflections.

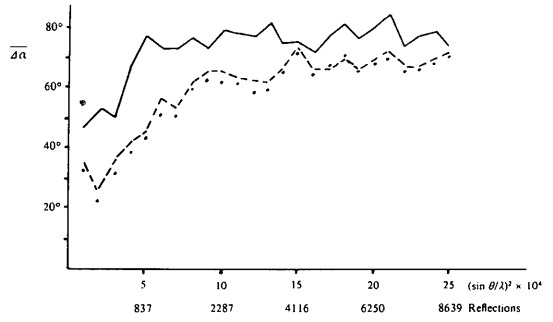

Fig. 3. Phase difference between the best 10 Å model, and the spherically averaged model (solid line), spherically averaged following seven cycles of phase refinement (broken line) and after eight cycles (dotted line).

Table 1. *Comparison of ab initio* $|F_c|$ *and measured* $|F_o|$ *values*

Resolution range (Å)	Number of reflections	R*	$\overline{\lvert F_c\rvert}/\overline{\lvert F_o\rvert}$
>15	167	0·19	1·04
15·0–12·0	222	0·20	0·97
12·0–10·0	510	0·17	1·00
10·0– 9·0	571	0·22	1·06
9·0– 8·0	980	0·25	1·06
8·0– 7·0	1927	0·39	0·92
7·0– 6·5	1748	0·46	0·90
6·5– 6·0	2655	0·53	0·81

* $R = 2\sum||F_o| - |F_c||/\sum(|F_o| + |F_c|)$.

Subsequently, another native data set recorded by the oscillation method to 4 Å resolution became available (Unge *et al.*, 1980). This afforded an opportunity to compare the 'pie slice' $|F_c|$ values of the 5·5 Å phase refinement with their $|F_o|$ values, heretofore unknown. Table 1 shows the results of this comparison following scaling of the two data sets based on reflections present as measured in both sets. The agreement is essentially perfect to about 8 Å resolution; the gradually worsening R value at higher resolution partly reflects a generally observed tendency of $|F_c|$ values to be less than $|F_o|$ near the resolution limit.

The combined data set to 4·0 Å resolution contained 148 158 reflections, of which 124 395 were measured. Starting at 7·0 Å resolution (31 196 reflections, 30 436 measured), we carried out a second series of phase refinement and stepwise phase extension. In this series the resolution limit for the Fourier-inversion generation of $F_c(\mathbf{h})$ was increased by $\Delta(\sin^2\theta/\lambda^2) = 0{\cdot}00025$ Å^{-2} per cycle. This introduced 2200 to 3500 new F_c values per cycle. The $|F_c|$ values of newly introduced reflections were employed in the F_o synthesis for *two* consecutive refinement cycles; that is, measured amplitudes were employed only for reflections whose $\sin^2\theta/\lambda^2$ value was at least 0·00050 Å^{-2} inside the current resolution limit. This phase extension and refinement progressed from 7·0 to 4·0 Å resolution in 42 cycles. At 4·0 Å several cycles with all measured data gave a final R of 0·215.

This phase-extension procedure may well have been unnecessarily cautious; in retrospect no indications were found to suggest that larger steps in $\sin^2\theta/\lambda^2$, *i.e.* fewer cycles, would have led to a different result.

The phase changes in a ten-cycle sequence of phase refinement and extension from 5·3 to 4·6 Å resolution are shown in Fig. 4. The phases for every other cycle are shown, relative to those of the last cycle in the sequence. The sharp rise at the high end of each curve reflects the uncertainty in the initial phases of the small, incipient F_c's.

Two likely contributing reasons for the relatively slow convergence of the phase refinement are the cycle-by-cycle updating of both the $|F_c|$ values of unmeasured reflections and of the solvent correction. A speed-up feature was tested whereby the structure factors $\mathbf{F}_c^{(n)}$ from the inversion of the symmetry-averaged map of cycle n were replaced by

$$\mathbf{F}_c^{(n-1)} + k[\mathbf{F}_c^{(n)} - \mathbf{F}_c^{(n-1)}]$$

with $k > 1$, when calculating input phases for the next cycle. The refinement progressed safely with values of k up to at least 1·5. In the refinements shown in Figs. 3 and 4 $k = 1$ was used.

The reliability of the phase extension is supported by the emergence of chemically interpretable features in the 4 Å electron-density map, including subunit boundaries and other significant features of the tertiary structure of the coat protein (Strandberg *et al.*, 1979; Unge *et al.*, 1980).

At 4·0 Å resolution, with 148 158 reflections, a map interval of 1·30 Å and a symmetry-grid step of 1·05 Å, the symmetry-averaging computation requires 4·5 min on the Amdahl 470V/7 computer. At this map density the stored STNV asymmetric unit requires 2·06 Mbytes of memory, at one byte per map point. The total memory requirement of this step is 2·5 Mbytes. An entire refinement cycle requires 18·6 or 13·0 min CPU time depending on whether or not the separate nucleic-acid Fourier-inversion step is included. The Fourier inversion and Fourier synthesis steps in the cycle (Ten Eyck, 1973) require 5·4 and 1·6 min of CPU time using buffered *I/O*.

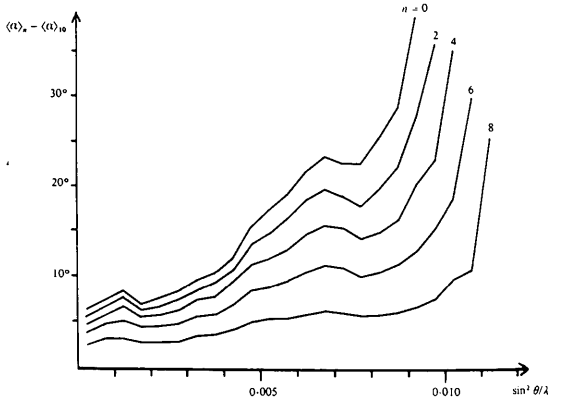

Fig. 4. Phase changes in ten cycles of phase refinement and extension from $(\sin\theta/\lambda)^2 = 0{\cdot}009$ to 0·012. Phase differences between cycles labelled $n = 0, 2, \ldots, 8$ and cycle 10, locally averaged in intervals of $(\sin\theta/\lambda)^2 = 0{\cdot}0005$, are shown as functions of $(\sin\theta/\lambda)^2$.

The author is indebted to Bror Strandberg, Uppsala, Sweden for stimulating cooperation and for making several sets of STNV X-ray data available for program development. Helpful discussions with P. J. Lentz Jr are gratefully acknowledged. The work was supported by grant GM 15259 from the National Institutes of Health.

References

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

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

BLOOMER, A. C., CHAMPNESS, J. N., BRICOGNE, G., STADEN, R. & KLUG, A. (1978). *Nature (London)*, **276**, 362–368.

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

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

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.

MOEWS, P. C. & KRETSINGER, R. H. (1975). *J. Mol. Biol.* **91**, 201–228.

NORDMAN, C. E. (1966). *Trans. Am. Crystallogr. Assoc.* **2**, 29–38.

ROSSMANN, M. G. (1972). *The Molecular Replacement Method.* New York: Gordon & Breach.

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

SCHILLING, J. W. (1970). *Crystallographic Computing*, edited by F. R. AHMED, pp. 115–123. Copenhagen: Munksgaard.

STRANDBERG, B., UNGE, T., LILJAS, L., VAARA, I., KANNAN, K. K., FRIDBORG, K., BORELL, A. & NORDMAN, C. E. (1979). *Am. Crystallogr. Assoc. Abstr. Ser.* 2, Vol. 7, p. 13.

STRANDBERG, B., VAARA, I., UNGE, T., FRIDBORG, K., KANNAN, K. K., BORELL, A., LENTZ, P. J. & NORDMAN, C. E. (1978). *Am. Crystallogr. Assoc. Abstr. Ser.* 2, Vol. 6, p. 29.

SUCK, D., RAYMENT, I., JOHNSON, J. E. & ROSSMANN, M. G. (1978). *Virology*, **85**, 187–197.

TEN EYCK, L. F. (1973). *Acta Cryst.* A**29**, 183–191.

UNGE, T., LILJAS, L., STRANDBERG, B., VAARA, I., KANNAN, K. K., FRIDBORG, K., NORDMAN, C. E. & LENTZ, P. J. (1980). *Nature (London)*. In the press.