Acta Cryst. (1983). A**39**, 102–116

Molecular Replacement Method at Low Resolution: Optimum Strategy and Intrinsic Limitations as Determined by Calculations on Icosahedral Virus Models

BY IVAN RAYMENT

Rosenstiel Basic Medical Sciences Research Center, Brandeis University, Waltham, Massachusetts 02254, *USA*

(*Received* 19 *April* 1982; *accepted* 28 *July* 1982)

Dedicated to Dr H. M. M. Shearer

Abstract

An extensive series of phase refinements on model structure factors has been carried out to determine the scope and limitations of real-space symmetry averaging at low resolution. These calculations were used to derive and justify the strategy used to solve the structure of polyoma virus capsids at 22·5 Å resolution [Rayment, Baker, Caspar & Murakami (1982). *Nature* (*London*), **295**, 110–115]. The calculations showed (1) that phases from a wide variety of models can be successfully refined; (2) a low R factor (~10%) between the constrained and observed amplitudes corresponds to a small error between the calculated and true phases; (3) on convergence to a low R factor there is no residual bias in the final phases introduced by the initial phasing model. The model calculations were also used to determine the limitations imposed by series termination, unmeasured data, random errors in the data and the role of the molecular envelope on the phase refinement.

Introduction

The crystal structure of mouse polyoma virus capsids has been solved at 22·5 Å resolution by X-ray diffraction (Rayment, Baker, Caspar & Murakami, 1982). The resultant electron density map shows that the organization of the viral capsid does not correspond to that expected from the principles of virus design and precepts of quasi-equivalence set forward by Caspar & Klug (1962). The surface of polyoma virus seen in electron micrographs of negatively stained particles is composed of 72 morphological units that are located at the 12 five- and 60 sixfold vertices of a $T = 7d$ icosahedral surface lattice (Finch, 1974). From this it was presumed that the virus capsid would be composed of 60 hexamers and 12 pentamers of the major coat protein. The crystal structure of the virus capsid shows, however, that all 72 morphological units are pentamers. Because this result contradicts a theory which has accommodated all previous structural knowledge of viral architecture, the method used to determine the structure has been critically evaluated.

The structure was solved using the fivefold noncrystallographic symmetry present in the diffraction data. The virus capsids crystallize in space group $I23$ where only the tetrahedral subgroup of the particle icosahedral symmetry is expressed in the crystal lattice. The fivefold noncrystallographic symmetry was used to refine, by symmetry averaging in real space, an initial set of phases derived from a model constructed according to the dimensions and surface features deduced from electron microscopy and small-angle X-ray diffraction.

The purpose of this paper is to demonstrate that the strategy adopted for solving the structure of polyoma virus capsid can yield the correct solution. This was accomplished by trying all the procedures that were used to refine the real capsid structure on a model set of amplitudes. The use of model amplitudes instead of observed data to test the method had the advantage that the correct phases were known. Thus, it was possible to judge the quality of the refinement procedures by comparison of the refined and correct phases. The questions addressed by the model calculations were: (*a*) What are the intrinsic limitations in the refinement process? (*b*) What are the effects of errors and unmeasured data on the refinement? (*c*) Can low-resolution phase refinement and extension distinguish between hexameric or pentameric morphological units? and (*d*) Does the initial phase set bias the final result?

The refinement method

The use of noncrystallographic symmetry in phase refinement is well established. It has played an essential role in the determination of four virus structures at high resolution: the protein disk of tobacco mosaic virus (Bloomer, Champness, Bricogne, Staden & Klug, 1978), tomato bushy stunt virus (Harrison, Olson,

0567-7394/83/010102-15$01.50

Schutt, Winkler & Bricogne, 1978), southern bean mosaic virus (Abad-Zapatero, Abdel-Meguid, Johnson, Leslie, Rayment, Rossmann, Suck & Tsukihara, 1980, 1981), and satellite tobacco necrosis virus (Unge, Liljas, Strandberg, Vaara, Kannan, Fridborg, Nordman & Lentz, 1980). These are currently the largest structures determined by X-ray diffraction analysis to molecular resolution.

Noncrystallographic symmetry was first proposed to be a source of phase information by Rossmann & Blow in 1962. Their expressions for applying the constraints operated in reciprocal space and were inconvenient to use since the number of computations increased with the square of the number of reflections. This problem was simplified by Bricogne (1974, 1976) who reformulated the expressions in terms of symmetry averaging in real space, as had been suggested by Main (1967) and Rossmann (1972). Symmetry averaging in real space has proved to be the computationally practical approach for using noncrystallographic symmetry as a phase constraint.

Phase refinement in real space is accomplished by Fourier transforming a symmetry-averaged electron density map. The initial electron density map is computed using the observed structure factors combined with phases derived from isomorphous replacement or from a previous cycle of phase refinement. The noncrystallographic symmetry is applied to the density within an envelope which contains the structure. Those parts of the map which are not included in the envelope are set to a constant value given by their average electron density. Refined phases obtained by Fourier transforming this modified electron density map form the basis for starting the next cycle of refinement. Programs for accomplishing these steps have been developed by Bricogne (1976), Johnson (1978) and Nordman (1980). The algorithm used in the work to be described follows the procedure of Johnson, which is essentially the same as that described by Bricogne.

Low-resolution phase refinement

The use of noncrystallographic symmetry is valid at low resolution; however, it is difficult to assess independently the quality of the refined phases. The correctness of the structures solved at high resolution has been judged by their stereochemical interpretability. This criterion is not applicable at low resolution.

The initial phases at very low resolution can be derived from measurement of the spherically averaged transform by small-angle X-ray scattering or by model building based on electron microscopy. These phases may not be randomly distributed about the correct phases and may consequently bias the final structure towards the initial presumptions. This is a particularly serious problem for phases generated from models because they constitute a self-consistent set which will differ systematically from the correct set. Thus, it must be demonstrated that the initial model does not impose an indelible imprint on the refined phases.

Two virus structures have been solved at low resolution. Tomato bushy stunt virus was solved initially at 28 Å resolution using information from solution scattering (Harrison, 1971). These phases were adequate to locate heavy-atom positions in an isomorphous derivative (Harrison & Jack, 1975). The resultant isomorphous phases were refined in reciprocal space to yield a low-resolution electron density map. Southern bean mosaic virus was solved at 22·5 Å by fitting the low-resolution diffraction data to the transform of a solid sphere (Johnson, Akimoto, Suck, Rayment & Rossmann, 1976). This approach was satisfactory because the surface morphology of the virus is quite smooth. The refined electron density map showed features that were confirmed at high resolution. In both of these structure determinations the initial phases were all centric. In the case of tomato bushy stunt virus the heavy-atom difference Fourier showed the unresolved average of the enantiomorphic images of the heavy-atom sites. The acentric locations were found using Fourier search methods. The enantiomorphic images of southern bean mosaic virus could not be separated. This imposed a limit on the resolution to which the *ab initio* phase determination could be taken. Introduction of the noncentric contribution to the initial or partially refined phases is important in order to obtain satisfactory phase refinement. For polyoma virus the $T = 7d$ distribution of morphological units in the starting model introduces the correct acentricity into the initial phases.

The numerical criterion by which to judge the convergence of the phase refinement is the R factor between the observed data and the calculated amplitudes from the constrained map. This value is a measure of how well the phases express the imposed noncrystallographic constraints. If the applied constraints are valid and stringent, the R factor should be a good measure of phase convergence. Thus, it is important to know what value of the R factor should be expected for a given set of experimental conditions before the refined phases can be accepted.

Phase extension

The molecular envelope and solvent flattening can be used as a constraint to derive phases to a higher resolution than is possible from the initial model calculations. The feasibility of phase extension was demonstrated by Argos, Ford & Rossmann (1975) at fairly high resolution on glyceraldehyde-3-phosphate dehydrogenase, which exhibits 222 molecular symmetry. Phase extension has also been tested by

extending phases from 10 to 4 Å resolution for satellite tobacco necrosis virus (Nordman, 1980). In both of these cases, the structure of the molecular aggregates had already been obtained by isomorphous replacement and refined by conventional molecular replacement so that the results from *ab initio* phase extension could be verified. At low resolution the method was used to phase reflections between 30 and 22·5 Å resolution for southern bean mosaic virus (Johnson, Akimoto, Suck, Rayment & Rossmann, 1976). All the structural features in the low-resolution electron density map were confirmed in the subsequent high-resolution structure determination.

The low-resolution electron density map for polyoma virus did not show the pentamer–hexamer arrangement of protein subunits expected from the precepts of quasi-equivalence. The model calculations to be described were undertaken to demonstrate that phase refinement and extension can yield definitive structural results at low resolution. A further objective was to establish the best strategy for utilizing the non-crystallographic constraints.

Creation of a model data set

Model diffraction data sets were calculated by Fourier transforming models which exhibited many of the features of polyoma virus capsid structure inferred from electron microscopy and small-angle diffraction. These data sets were used to characterize the expectations and limitations of the molecular replacement method at low resolution. The findings were reasonably independent of the model used to generate the structure factors, although the exact statistics varied. The experiments described here were performed using structure factors from one basic model so that the relative effects of changing a given parameter could be seen. The amplitudes and phases of these structure factors, which took the place of the observed polyoma diffraction data, are designated the 'native' model amplitudes and phases throughout the rest of the paper.

The amplitudes and phases were generated by Fourier transforming a model electron density map whose unit-cell dimensions and space group were identical to those of the polyoma capsid crystals ($I23$, $a = 572$ Å). The strategy for calculating the maps is summarized in Table 1.

Table 1. *The strategy for building a model electron density map*

1. Calculate the locations of the structural entities in asymmetric unit.
2. Determine which grid points are associated with the structural entity.
3. Calculate the density value for each point.
4. Pack the coordinates of each point into one word.
5. Store the packed word and density value.
6. Sort the entire list of grid points on section number.
7. Construct the electron density map.
8. Fourier transform the map to obtain structure factors.

The image reconstruction of polyoma virus showed that the surface consists of 72 stubby protrusions with no discernible substructure, arranged on a $T = 7d$ icosahedral surface lattice (Finch, 1974). These features were modeled by creating a map in which the virus capsid was built from 72 hollow cylinders 42 Å tall with an external and internal diameter of 70 and 34 Å, respectively, representing the protruding domains. These morphological units rested on a shell of density extending from a radius of 190 to 200 Å. The surfaces of the model were softened by adding a Gaussian fall-off of 2 Å width at half-height, extending the radius of the capsomers to ~247 Å before trailing off. The phases from this model were adequate for initiating the refinement against the real capsid data. Additional features were added to this simple model in order to generate a more realistic set of structure factors for testing the refinement procedures. The density of the inner shell was set to twice that of the protruding cylinders. Then substructure was added to the model by embedding small cylinders of density 12 Å in diameter into the walls of the morphological units. Five small cylinders were placed in each of the 12 pentavalent units whereas six small cylinders were placed in each of the 60 hexavalent units. In this way the substructure showed a $T = 7$ distribution in which the small cylinders represented the expected 420 quasi-equivalently related protein subunits. The density of the small cylinder subunits was equal to that of the hollow cylinders in which they were embedded so that together their maximum density was never greater than that of the inner shell. The small cylinder substructure constituted 10% of the total density in the model. This small contribution was chosen to test the sensitivity of the refinement method and to mimic the substructural density fluctuations observed in the electron density maps of the polyoma capsomers.

The model was further elaborated by placing 60 holes in the shell, 34 Å in diameter, located close to one of the local threefold axes. These were added because similar density fluctuations were seen in the real polyoma capsid density map and it was of interest to see if the refinement could return these features in the computed model maps. A projection of this model electron density map at 22·5 Å resolution down an icosahedral fivefold axis and through a hexavalent unit is shown in Figs. 1(*a*) and (*b*).

Intrinsic limitations in the refinement

It is important to establish if there are any features of the refinement method and algorithm that will prevent the phases from converging to their correct values. This

is necessary in order to establish a standard by which the refinements may be judged. The 'native' model amplitudes were free from errors of measurement and therefore represent the best possible data for testing the method.

There are two intrinsic limitations: firstly, series termination and, secondly, inaccuracy of the averaging algorithm. Series termination has the greatest effect. This results in diffraction ripples which extend throughout the solvent and density regions of the map. The magnitude of the ripples increases as the resolution of truncation decreases. Thus, any electron density map calculated from correct amplitudes and phases will not have flat solvent regions unless all the data are included in the calculation.

The refinement process entails imposing the non-crystallographic symmetry on the electron density map and flattening the solvent regions. New phases and calculated amplitudes are obtained by Fourier transforming this constrained map. The phases within the resolution limit will adjust themselves to produce an artificially flat solvent region if the data have been truncated. Consequently, the refined phases will differ from their correct values as will the calculated amplitudes. This effect is more pronounced for truncation at 30 Å than at 22·5 Å resolution.

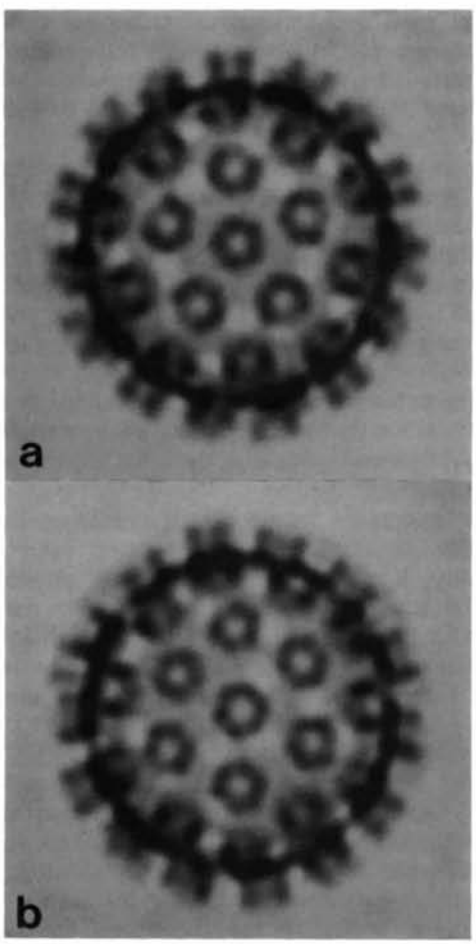


Fig. 1. Density maps of half of the native model at 22·5 Å resolution projected down (*a*) the icosahedral fivefold axis and (*b*) through the hexavalent morphological unit.

The effect of solvent flattening was shown by calculating an electron density map using the 30 Å structure factors and perfect phases from the 'native' model. The solvent regions were set to their average value leaving the electron density within the envelope untouched. The modified map was Fourier transformed to give a new set of calculated amplitudes and phases. The R factor between the calculated and original structure factors was 4%, where R is defined as

$$R = \frac{\sum ||F_m| - |F_c||}{\sum |F_m|} \times 100,$$

where $|F_c|$ and $|F_m|$ are the amplitudes for the modified and true 'native' model structure factors. The r.m.s. (root mean square) difference between the calculated and true phases was 15°. The data most affected by the solvent flattening were those structure factors on the edge of the resolution range. Between 30 and 31 Å resolution the R factor was 28% with an r.m.s. phase change of 43°.

The diffraction ripples also extend through the density regions of the map. The ripples contain a fivefold component which is not lost on averaging. The magnitude of the effect was determined by averaging the same 30 Å resolution electron density map calculated previously but leaving the solvent regions unchanged. The reconstructed map was again Fourier transformed to give calculated amplitudes and phases. The R factor between the initial and calculated amplitudes was ~2% with a 15° r.m.s. phase change. This calculation includes the errors introduced by the numerical interpolation of the averaging algorithm.

The errors associated with the numerical evaluation of the symmetry-related points in the electron density map were found by *building* a map which exhibited true noncrystallographic symmetry. This model map contained no truncation ripples. The electron density within the envelope was averaged and placed back into the original map. Both the original and averaged map were Fourier transformed to give original and modified structure factors. The differences between these data sets arise only from the errors introduced by the interpolation algorithm used for deriving density values at non-integral locations in the electron density map. This problem has been discussed in detail by Bricogne (1976). The magnitude of the error is in part dependent upon the fineness of the sampling interval at which the original map was calculated. An R factor of ~1% with an r.m.s. phase change of ~7° was obtained when the map was calculated at intervals of one-fifth of the

spacing of the highest-resolution data included in the calculation. The error was larger if a coarser grid was used. In all the subsequent calculations, a sampling of one-fifth of the resolution limit, or finer, was used for the initial map calculation.

The combination of both averaging and solvent flattening with this model data set at 30 Å resolution results in a 6·5% change in the amplitudes and r.m.s. phase change of 25°. These calculations show that the major limitation on convergence is caused by series termination. They also demonstrate that phases refined by noncrystallographic averaging and solvent flattening cannot converge on an absolutely correct set if the data are truncated. For this model data set at 30 Å resolution an R factor of better than 6·5% between the calculated and true amplitudes cannot be obtained by real-space refinement even when starting from the true phases.

Limitations imposed by unmeasured data

The absence of data within the resolution sphere of the refinement imposes a serious limitation on the phase convergence. This problem has been discussed previously by Nordman (1980) in the refinement and phase extension for satellite tobacco necrosis virus. Nordman demonstrated that inclusion of calculated values for the missing data improved the self-consistency of the phase extension considerably. Furthermore, when the missing data were eventually measured, comparison of their experimental and calculated values showed a very consistent agreement. In order to assess the effect of unmeasured data at low resolution a series of calculations were undertaken using 90% of the model data.

The data set collected from crystals of polyoma virus capsids included measurements for 94% of the data out to 22·5 Å spacing. In particular, no terms were recorded below 150 Å resolution. These conditions were simulated by removing a random 10% of the 'native' model data between 30 and 200 Å resolution together with all the terms with spacings larger than 200 Å. The intrinsic limitation imposed by the missing data was shown by calculating a map using the remaining correct amplitudes and phases. The density regions of this map were symmetry averaged over the noncrystallographic fivefold axis and the solvent regions were given their average value. The modified map was Fourier inverted to give a new set of calculated amplitudes and phases. The overall R factor between the constrained and native model data for the 90% included in the calculation was 21% with an r.m.s. phase change of 33°. This should be compared to the R factor and r.m.s. phase change of 6·5% and 25° when the noncrystallographic constraints were applied to a map calculated from 100% of the data out to 30 Å spacing. The absence of terms with spacings greater than 200 Å together with a random 10% between 30 and 200 Å spacing introduces significant changes in the structure factors derived from the constrained map. This effect as shown later will prevent satisfactory convergence of phase refinement and extension.

Phase refinement by symmetry averaging depends upon the continuity of the Fourier transform in reciprocal space. Ignoring unmeasured data introduces artefactual discontinuities in the transform. The refinement process smooths out the perturbations reducing the value for the terms close to the missing data and predicting an amplitude and phase for the missing data.

The absence of the low-resolution terms generates low-frequency ripples which propagate throughout the map. The 10% of the data missing between 30 and 200 Å resolution introduces higher frequency ripples that perturb the fivefold symmetry in the density regions of the map. Averaging the map and flattening the solvent regions smooths out the perturbations and generates calculated amplitudes and phases for the missing data. These calculated values for the missing data show a good agreement with their actual values. The overall R factor for the missing data between the calculated and true amplitudes was 21% with an r.m.s. phase difference of 29°.

The restrictions introduced by unmeasured data may be overcome in two ways. The best solution would be to collect 100% of the data; however, this is not usually practical. Alternatively, the effect may be mitigated by including the calculated values for the absent data derived from the previous cycle. When the calculated amplitudes and phases for the 10% missing data just described were included in four additional cycles of refinement, the R factor between the refined and native amplitudes for the 'observed' 90% of the data fell from 21 to 8% with a decrease in the r.m.s. phase difference of 33 to 27°.

The calculated values for the missing data are generally smaller than their correct values. Consequently, it was found acceptable to include these terms at a fairly high weight of ~0·9. The calculated values for these terms improved considerably in further cycles of refinement after their inclusion. The overall R factor between their actual and calculated amplitudes fell to 7·4% with an r.m.s. phase difference of 15° after four cycles.

In practical terms the inclusions of estimated values for any unmeasured reflections requires that the structure-factor list contain an entry for all of the recorded data whether it is above background or not. The zero-intensity data points in the list then explicitly specify the nodes in the transform. In this way calculated values will only be included for those points which are unrecorded, while measured nodes remain zero.

Tests on convergence

The major purpose of the model calculations was to demonstrate that the refinement of a trial set of phases against an 'observed' set of amplitudes using molecular replacement can lead to the true set of phases belonging to the amplitudes. An important aspect of these model studies was to determine to what extent the initial phase set influences the final solution. In order to assess the effect of the initial phases, a trial set of phases was generated by Fourier transforming a map whose density distribution differed from the 'native' model in a variety of ways.

The trial phases were combined with the native model structure-factor amplitudes and used to calculate an electron density map. The structure factors were weighted according to the fit of calculated amplitudes of the phasing model to the native model amplitudes. The weighting algorithm used took the form

$$w = e - \frac{||F_c| - |F_m||}{|F_m|}.$$

This simple weighting scheme, which will be discussed later, proved to be essential when the phasing model differed considerably from the native model.

Numerous sets of phases were tested. These may be placed into three classes: (*a*) phases from reasonable models built using the information that would be available from electron microscopy, solution scattering and physico-chemical data; (*b*) phases from a rudimentary model; and (*c*) phases from a misleading model which contains information that is incorrect, in an attempt to bias the final refined phases.

Phases from all three classes of models refined towards the true set of phases within the limitations imposed by series termination. The number of cycles of refinement necessary for convergence was proportional to the magnitude of the initial phase error. These experiments showed that appropriately constrained phase refinement can lead to convergence to a phase set whose corresponding electron density map was indistinguishable from the map calculated from the real phases.

Refinement of phases from a reasonable model

The 'reasonable' phase model that was chosen consisted of 72 hollow cylinders of density 70 Å in diameter with a 34 Å hole extending from 200–235 Å located on a $T = 7d$ surface lattice. These cylinders rested on a concentric shell of density extending from 180 to 200 Å. The location of the hexavalent unit was ~10 Å from that used to generate the native model. The major differences between this phasing model and the native model structure were the radial extent of the concentric shell (180–200 *vs* 190–200 Å) and morphological units (200–235 *vs* 200–242 Å) together with the lack of substructure. This model is considered reasonable because it was built according to the information that would have been available from electron microscopy, solution scattering and crystal packing considerations. There were no presumptions made about the internal structure of the native model.

The phases from this reasonable model were combined with the amplitudes from the native model out to 30 Å resolution and used to calculate an electron density map. This map was symmetry averaged as described before and Fourier inverted to obtain a new set of improved phases. This process was repeated using the phases from the last cycle to calculate the initial map of the next cycle until there was no further improvement in the R factor between the calculated and native model structure factor amplitudes.

The statistical results of this process are summarized as a function of resolution and refinement cycle in Fig. 2. The initial R factor between the amplitudes of the phase model and native model was 40% with an r.m.s. phase difference of 72°. The R factor after 15 cycles was 7·1% with an r.m.s. phase difference of 27°. The major error lies at the edge of the resolution range where the series termination has its major effect. Fig. 2(*c*) shows the overall R factor as a function of cycle number. The major progress in the refinement occurs in the first four cycles. Thereafter the refinement converges slowly.

The refinement of model phases occurs more slowly than the refinement of phases from isomorphous replacement. Model phases usually require 10–20 cycles of symmetry averaging before convergence is reached, in contrast to the refinement of phases derived from multiple isomorphous replacement (MIR) which is usually complete in 3–5 cycles. This difference in behavior of MIR and model phases arises because of the nature of the reliable phase information present in each type. The phase information in MIR phases is more reliable at moderate resolution than at low resolution and is distributed about the correct set of phases. In contrast, phases from a model are more accurate at low resolution and contain no details of the high-resolution structure. Furthermore, the model phases constitute a self-consistent set whose corresponding density differs in a systematic way from the true electron density distribution. Consequently, it takes longer to remove the initial bias resulting in slow convergence.

Refinement of phases from a rudimentary model

The previous experiment demonstrates that phase refinement certainly works although this test did not

pose any serious difficulties. A more significant test is to see how little information is required in the phasing model in order to initiate phase refinement successfully. This test might also illustrate if an alternative phase solution corresponding to a significantly different electron density distribution could be consistent with the constraints and the native model structure factors.

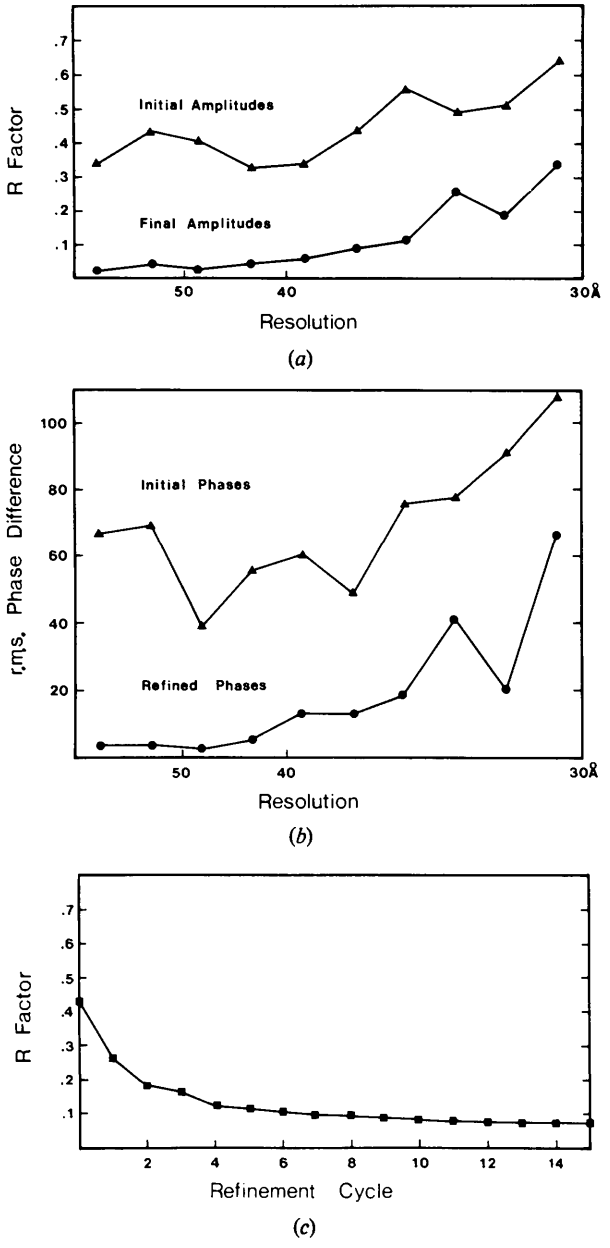


Fig. 2. Statistics for the refinement of phases from the reasonable model against the native model amplitudes at 30 Å resolution. (*a*) The *R* factor as a function of resolution for the initial and final amplitudes; (*b*) the r.m.s. phase difference for the initial and final phases; and (*c*) the *R* factor as a function of refinement cycle.

The model chosen for this test consisted of 60 spherical morphological units 70 Å in diameter placed close to the location of the hexavalent axes of a $T = 7d$ surface lattice. The centers of the spheres were placed 205 Å from the center of the particle. These spheres intersected a concentric shell of density extending from a radius of 180–200 Å. No density was placed on the 12 pentavalent locations. Although this model is rudimentary, the basic presumption that the particle has icosahedral symmetry is still maintained. Icosahedral symmetry was presumed for all models.

The phases from the rudimentary model were combined with the amplitudes from the native model and used to calculate an electron density map. The initial *R* factor between the two sets of amplitudes was 55% with an r.m.s. difference of 93° between the phases. An illustration of the initial phase combination prior to averaging and solvent flattening is shown in Fig. 3(*c*). The map section corresponds to a plane perpendicular to the axis of a hexavalent unit 220 Å from the center of the model. The cross section of the unaveraged map shows that the initial combination is very noisy. Comparison of this initial map (Fig. 3*c*) with the same section from the native model (Fig. 3*a*) and phase model (Fig. 3*b*) shows clearly that the initial phase dominate the first density map.

Refinement of this set of phases proceeded far more slowly than those from the reasonable model described previously. After 15 cycles of refinement the *R* factor was 11% with an r.m.s. phase difference from the true set of 51°. With an additional six cycles of refinement the *R* factor dropped to its final value of 7·2% with an r.m.s. phase difference of 28°. The major reason for the slowness of the refinement is the large number of incorrect signs amongst the low-resolution centric reflections. Initially 42% of the centric reflections had the wrong sign. It is difficult for a centric reflection to change sign unless some form of weighting scheme dependent upon the fit of the individual observed to calculated structure factors is used. Consequently, it takes several cycles of down-weighting for a strong centric reflection with the incorrect sign to have its sign switched by its noncrystallographically related neighbors with the correct signs.

A section through the final refined map is shown in Fig. 3(*d*). This is virtually indistinguishable from the correct 30 Å native model map shown in Fig. 3(*a*). The calculations illustrated in Fig. 3 demonstrate that the weighted phase refinement with accurate data does converge on the correct phase solution. This example is particularly interesting because the phasing model had no density on the fivefold axes. The refinement constraints and weighting procedures allowed the phases to change so that the information in the amplitudes could be expressed. This shows that the refinement method can return structural information for features that were not included in the phasing model.

Refinement of phases from a misleading model

An important test of the method is to see if it is possible to bias the final refined phases by introducing incorrect information in the starting phases. This is a significant question, since, in the real polyoma capsid data, all 72 of the morphological units show fivefold substructure. It is important to determine if the fivefold nature of the hexavalent unit could be caused by systematic errors in the starting phases.

This question was approached by taking the phases from a model whose hexavalent units contained a strong fivefold substructure. All the morphological units were composed of five cylinders of density with the same weight as the concentric shell of data upon which they rested. The all-pentamer substructure constituted 30% of the density in the phasing model as compared to 10% in the native model that was built with hexameric hexavalent units. One section of the enhanced fivefold phasing model is shown in Fig. 4(*b*). The combination of the phases out to 30 Å resolution with the amplitudes of the native model is shown in Fig. 4(*c*). In this example the fivefold character of the misleading phasing model is expressed in the initial electron density map. The R factor between the amplitudes of the two data sets was 47% with an r.m.s. phase difference of 71°.

After 15 cycles of refinement the R factor between the calculated and refined amplitudes was 6·9% with an r.m.s. phase difference of 26°. This phase set has converged to the expected limits imposed by series termination. The resultant electron density map (Fig. 4*d*) is indistinguishable from the true 30 Å resolution map (Fig. 4*a*). This refinement shows that there is no residual bias introduced by the misleading starting model in the final phase set once convergence to the limits imposed by truncation has occurred.

The influence of unmeasured data on phase convergence

So far the refinement tests have been performed using *all* the data within the 30 Å resolution sphere. As noted previously, this is an unrealistic situation. The unmeasured data increase the R factor between the

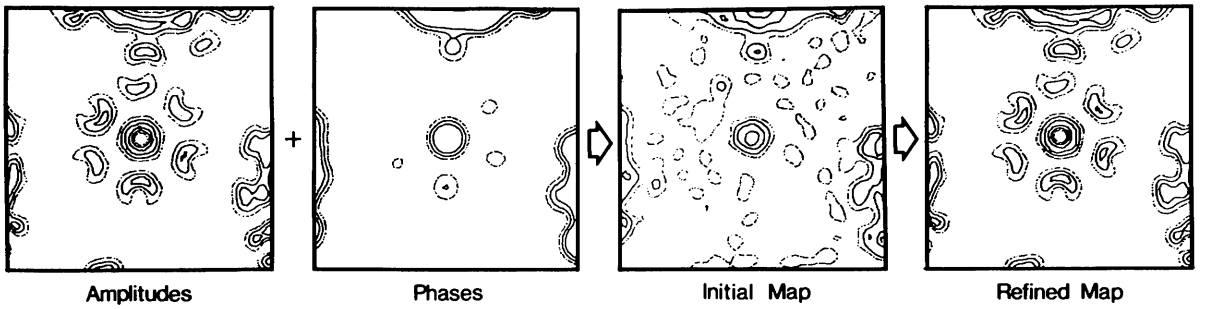


Fig. 3. Refinement of phases from the rudimentary model at 30 Å resolution. The maps are perpendicular to the axis of a hexavalent morphological unit, sectioned 220 Å from the particle center. (*a*) The native model; (*b*) rudimentary phasing model; (*c*) combination of the phases from the rudimentary model and native model amplitudes and (*d*) combination of the final refined phases with native model amplitudes.

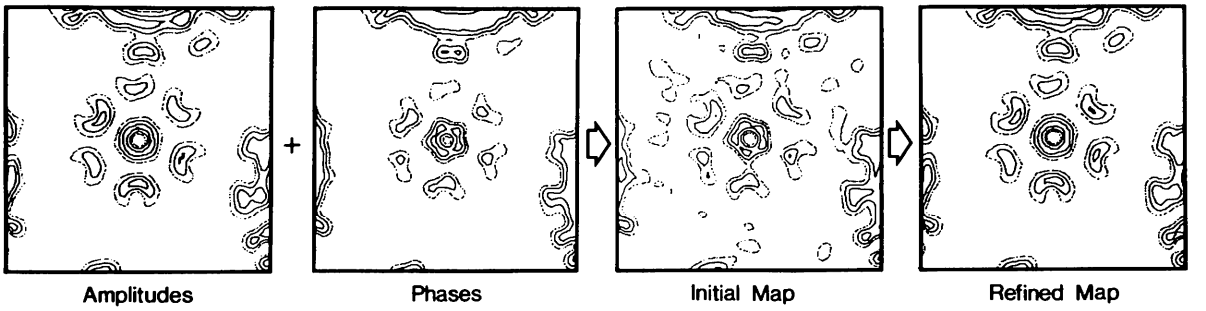


Fig. 4. Refinement of phases from a misleading model at 30 A resolution. The maps are perpendicular to the axis of a hexavalent morphological unit sectioned 220 Å from the particle center. (*a*) The native model; (*b*) misleading phasing model; (*c*) combination of the phases from misleading model and native model amplitudes; and (*d*) combination of the final refined phases with the native model amplitudes.

constrained and observed amplitudes but more seriously missing data can prevent satisfactory convergence of an incorrect set of phases, and can confuse the resultant density map.

The effect of unmeasured data on phase refinement was demonstrated by using the partial data set described previously. This contained 90% of the terms between 30 and 200 Å resolution, excluding all reflections with spacings greater than 200 Å. This incomplete data set was combined with phases from the reasonable model described previously and refined to convergence. In this case the final R factor between the calculated and true amplitudes was 24% with an r.m.s. phase difference of 48° between the true and refined phases. Convergence to these results occurs in about three cycles with no further improvement in the interpretability of the density map over an additional 15 cycles of refinement. The R factor and r.m.s. phase difference after 18 cycles as a function of resolution are shown in Figs. 5(*a*) and (*b*), respectively.

A section of the initial electron density map computed using phases from the reasonable model combined with the partial native model data set is shown in Fig. 6(*a*). The same section after 15 cycles of refinement but prior to the final averaging and solvent flattening is shown in Fig. 6(*b*). This map is very noisy. Averaging this map (Fig. 6*c*) significantly reduces the noise as would be expected. However, there are still considerable changes in the electron density distribution introduced by the missing data (*cf.* Fig. 4*a*). Even at 30 Å resolution the absence of 10% of the data significantly confuses the substructure in the hexavalent morphological units.

The previous refinement experiment was repeated with the inclusion of the calculated values for the missing data starting after the second cycle. The missing terms were added in shells of increasing resolution over four cycles. After 15 cycles of refinement the R factor for the 'measured' 90% was 6·7% with an r.m.s. phase difference of 28°. The distribution of R factor and phase difference for these data is also shown in Figs. 5(*a*) and (*b*). The resultant electron density map (Fig. 6*d*) is indistinguishable from the

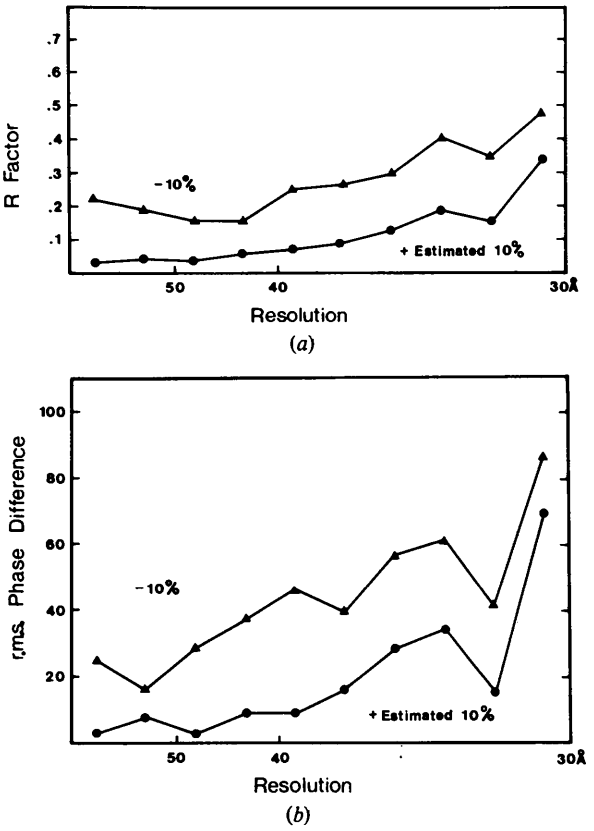


Fig. 5. The distribution of (*a*) R factors, and (*b*) r.m.s. phase differences for refinement after 18 cycles of refinement with 10% of the data removed and all terms with spacings larger than 200 Å (▲). Inclusion of the calculated values for the missing data (●) gives a significant improvement in the quality of the refinement.

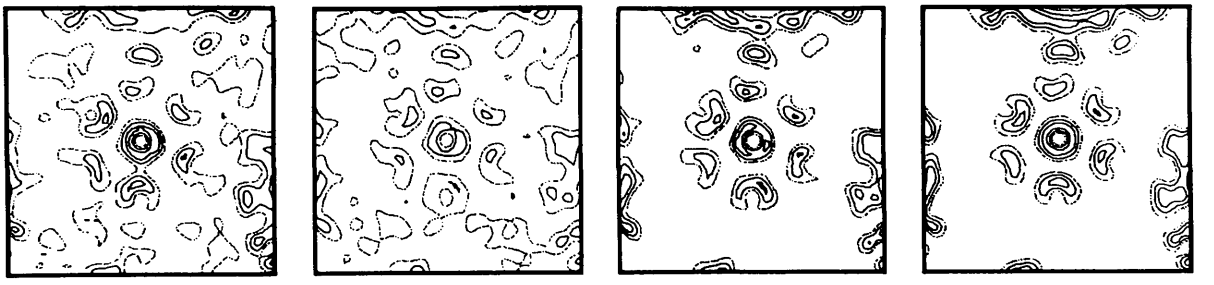


Fig. 6. The effect of unmeasured data on the phase refinement at 30 Å resolution. (*a*) Initial combination phases from the reasonable model with 90% of the native model amplitudes; (*b*) unaveraged and (*c*) averaged maps calculated using phases refined using 90% of the data; (*d*) the unaveraged map calculated from phases refined including calculated values for the missing 10% of the data; this is virtually indistinguishable from the same section of the true native model shown in Fig. 3(*a*).

correct map (Fig. 3*a*). The comparison of the sections with and without the calculated values for the 'unmeasured' data (Figs. 6*c* and *d*, respectively) demonstrates the importance and necessity of compensating for this data.

The *R* factor and r.m.s. phase difference between the calculated and true values for the 'unmeasured' data were 6% and 17°. The agreement between the calculated and true values for the data that were left out is slightly better than that obtained for the 90% of the data that were included as 'observed'. This occurs because the *R* factor is dominated by the strong low-resolution terms which are accurately regenerated by the solvent flattening. The higher-resolution missing data can be recovered because the noncrystallographic fivefold symmetry leads to an oversampling of the particle transform. Averaging density maps regenerates the continuous transform of the icosahedral particle.

The role of the envelope

The purpose of the envelope is to define which parts of the unit cell are to be symmetry averaged. All other points are treated as solvent and set to some uniform value. It is important to contain all the density within the envelope. Failure to enclose all the density will result in a large *R* factor between calculated and observed data as shown later.

Evidence for an incorrect envelope can be seen by looking at an averaged electron density map. The gradient across the envelope boundary will be a smooth function except where the density has been incorrectly truncated. In practice this requires that the envelope should be looser than that suggested by molecular weight and partial specific volume measurements. It must be remembered that at low resolution the boundaries between protein and solvent are not well resolved. Thus, the density will extend beyond any volumes dictated by physical-chemical information.

The solvent regions place a constraint on the phases. The larger the volume that can be associated with solvent regions, the more powerful the constraint. Thus, the choice of envelope is a compromise between the necessity to enclose all the protein and yet leave as large a volume as possible in the solvent regions.

There are two other considerations. Firstly, the envelope must show the particle symmetry. In the case of spherical viruses, the simplest envelopes are either spherical or icosahedral shells. More detailed envelopes, based on information of the surface structure may, if incorrect, prevent the refinement from converging. Secondly, the envelopes of adjacent particles may touch but not overlap. A point inside an envelope may belong to only one particle. This is a necessary requirement in order to locate correctly the noncrystallographically equivalent densities in order to generate the average for that point.

The role of the envelope was determined by refining the same phase sets with different envelopes. Refinements were performed using the phases from both the reasonable and rudimentary phase models described earlier.

The envelope used consisted of a concentric shell truncated by planes perpendicular to the icosahedral threefold axes. The virus particles are situated at (0,0,0) and $(\frac{1}{2},\frac{1}{2},\frac{1}{2})$ in the unit cell such that they contact one another along the body diagonals which are the crystallographic threefold axes. The three hexavalent units which lie adjacent to the crystallographic threefold axes contact one another tip-to-tip because of the way the hexavalent vertices are arranged on the $T = 7d$ icosahedral surface lattice. Fig. 7 shows a section through the native model virus particle, together with the envelope that was used to contain the density regions of the map. This envelope is the simplest icosahedrally symmetric non-overlapping envelope that can be packed in an $I23$ unit cell. The use of a spherical envelope would have truncated the density on the hexavalent and pentavalent morphological units.

The results of eight refinements at 30 Å resolution using four different envelopes are summarized in Table 2. The first envelope consisted of a concentric shell extending radially from 173 to 235 Å which truncated the outer 15 Å of the native model density. Envelopes 2, 3 and 4 were used to show the effect on the phase refinement of increasing the volume of the map which was symmetry averaged. These statistics show that the best phase refinement for both trial phase sets is obtained using a tight envelope (2), even though it results in slightly higher *R* factors than the looser envelope (3).

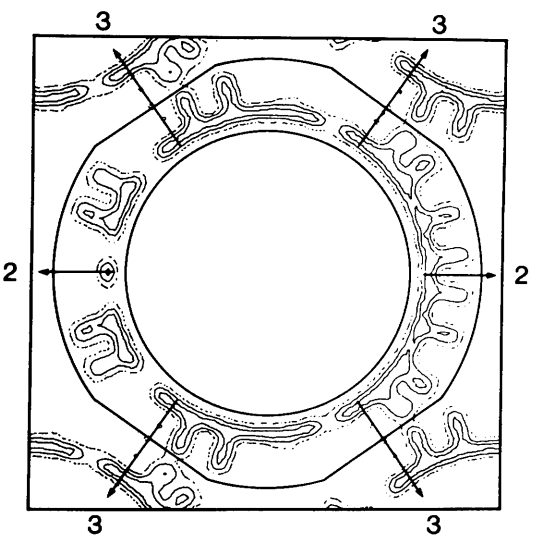


Fig. 7. A section perpendicular to the [110] direction through the middle of the native model showing the envelope used for defining the boundaries of the particle. This map shows that hexavalent units on adjacent particles contact each other tip-to-tip.

Table 2. *The effect of the choice of molecular envelope on phase refinement*

	External* radius (Å)	Internal radius (Å)	% Solvent†	Reasonable model R	Reasonable model r.m.s. $\Delta\phi$ (°)	Rudimentary model R	Rudimentary model r.m.s. $\Delta\phi$ (°)
1.	235	173	92·0	25·0	58	26·0	71
2.	252	173	52·0	7·1	27	7·2	28
3.	265	160	40·7	6·4	32	12·2	47
4.	265	0	22·3	7·9	47	16·0	78

* Envelopes 2, 3 and 4 are truncated by planes perpendicular to the 20 icosahedral threefold axes at a distance 247 Å from the center.
† Fraction of the unit-cell volume outside the envelope.

The importance of solvent flattening as a constraint in the phase refinement is shown by envelope 4 in which the interior of the particles was icosahedrally averaged but not set to the average solvent density. In this case the phases from the reasonable model converged to a set which gave a low R factor but poorer phases. When this envelope was used in the refinement of the phases from the rudimentary model the convergence was not as satisfactory. The solvent flattening makes a significant contribution towards changing the signs of incorrect centric reflections. This accounts for the inability of envelope 4 to switch the incorrect signs of the centric reflections of the rudimentary phase set.

The first envelope shows the effect of excluding part of the density, by truncating the outer 15 Å of the particle. This represents an exclusion of 17% of the density. Its absence severely restricts the convergence of the refinement as indicated both by the high R factors and r.m.s. phase differences.

Phase extension

Phase extension was necessary in solving the structure of polyoma virus because initial phases generated from models were not reliable beyond a resolution of ~30 Å. Attempts to refine model phases against the real data for all reflections out to 22·5 Å resolution were unsuccessful. Unsatisfactory convergence of the phases for the higher-resolution data results because it is not possible to model accurately the details of the density distribution within the morphological units or the inner shell. The diffraction to ~30 Å resolution is dominated by the volume and location of the morphological units. These parameters can be confidently modeled and result in a good set of phases for initiating the refinement at 30 Å resolution. Because approximately two thirds of the data were phased by extension, it was vital to demonstrate that phase extension is applicable at low resolution. This was accomplished using the model amplitudes.

Solvent flattening or application of the envelope to an electron density map is the source of the phases generated just outside the resolution sphere of data used in the map calculation. This was demonstrated by setting the solvent regions of an electron density map calculated using terms out to 30 Å resolution to their mean value without symmetry averaging. The modified map was Fourier transformed to yield new amplitudes and phases. As mentioned earlier, this introduces a 4% change in the amplitudes and 15° r.m.s. change in the phases for those terms included in the initial map calculation. In addition, the solvent flattening generates amplitudes and phases for reflections with spacings smaller than 30 Å. The R factor for these generated amplitudes compared with their true values for the data between 28·5 and 30 Å resolution was 37% with an r.m.s. phase difference of 47°. The phase difference for the most intense half of the data in this narrow shell was 25°, which demonstrates that phases can be reliably extended if the phases of the lower-resolution data are known.

The phase extension strategy adopted consisted of taking the phases for a narrow band of structure factors just outside the resolution sphere of the previous cycle of refinement and including them with their corresponding 'observed' amplitudes in the next cycle. These additional phases were refined for two cycles before any additional terms were added. In the first cycle the new reflections were given unitary weights; in the second and subsequent cycles they were weighted according to the fit of the calculated to the observed amplitudes. Additional shells of data were included in alternate cycles of refinement until all the data had been included.

Phase extension utilizing the strategy outlined above was tested on the native model data set. The initial phases at 30 Å resolution were obtained from the refinement of phases from the reasonable model described earlier. After 24 cycles of refinement and phase extension the overall R factor between the calculated and true native model amplitudes for all the data to 22·5 Å resolution was 6% with an r.m.s. phase difference of 32°. One section through the phase-extended electron density map is shown in Fig. 8(*b*). This should be compared with the same section through the native model map shown in Fig. 8(*c*). Even though there is a slight loss of detail, there is no ambiguity in distribution of the substructure. Comparison of the extended and refined electron density map with the original 30 Å map shows that the extension strategy

can yield correct structural details not present in the initial phases.

The quality of the phase extension and refinement decreased with increasing resolution. This is, in part, a consequence of the fall-off in intensity of the data but also a reflection of the weaker constraint of the envelope. A more detailed envelope at this stage would have yielded better phase prediction and constraint.

The effect of unmeasured data on phase extension

The absence of a small percentage of the data severely restricts the progress of phase extension and associated phase refinement. It was found that inclusion of calculated values for the 'unmeasured' structure factors and phases allowed the calculations to converge to an acceptable solution. The strategy adopted consisted of including calculated values for the unmeasured structure factors two cycles out of phase with the inclusion of the new phases for the observed data. This lag enabled the shell of data just included to refine itself and predict the values for the data which were absent.

Fig. 9(*a*) shows the refinement statistics as a function of resolution for phase extension with and without calculated values for the 'unmeasured' data. Inclusion of the calculated values for the absent data gave an overall *R* factor of 8·2% with a 35° r.m.s. phase difference compared to the true values; whereas their absence resulted in an *R* factor of 25% and 72° r.m.s. phase difference.

One section through a hexavalent unit for the averaged map after phase extension using 90% of the data is shown in Fig. 10(*a*). The same section is shown in Fig. 10(*b*) for phase extension after including calculated values for the 10% of the data which was considered 'unmeasured'. This map was not averaged. Comparison of the two sections with the true 22·5 Å map shown in Fig. 10(*c*) shows that failure to compensate for the unmeasured data will yield an incorrect and confusing electron density map.

The effect of errors in the data on real-space refinement

In all of the experiments described so far, the structure-factor amplitudes of the native model data set have contained no errors. In order to see the effect of noise in the data a random set of errors was imposed on

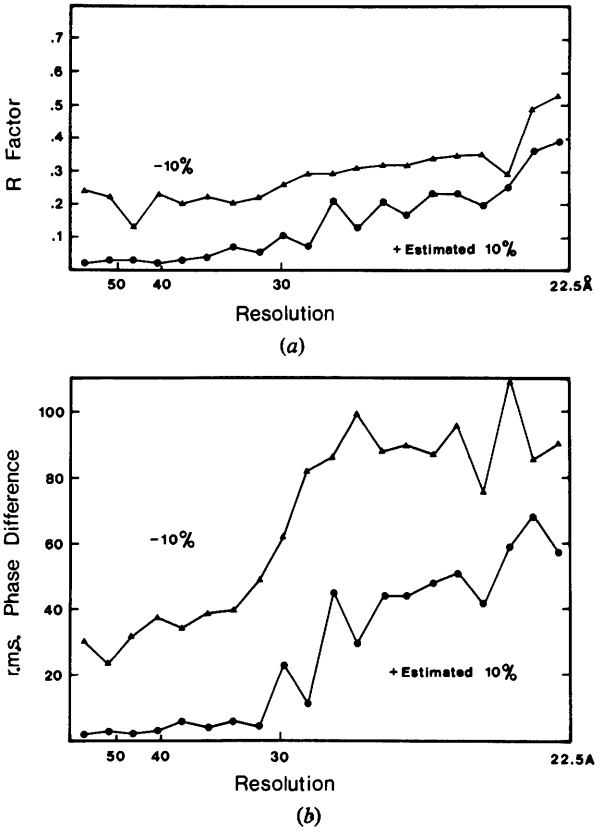


Fig. 9. The distribution of (*a*) *R* factor and (*b*) r.m.s. phase difference for phase extension and refinement including data with spacings larger than 200 Å but omitting a randomly distributed 10% of all of the higher-resolution data (▲). Inclusion of the calculated values for the missing data (●) gives a significant improvement in the quality of the phase extension.

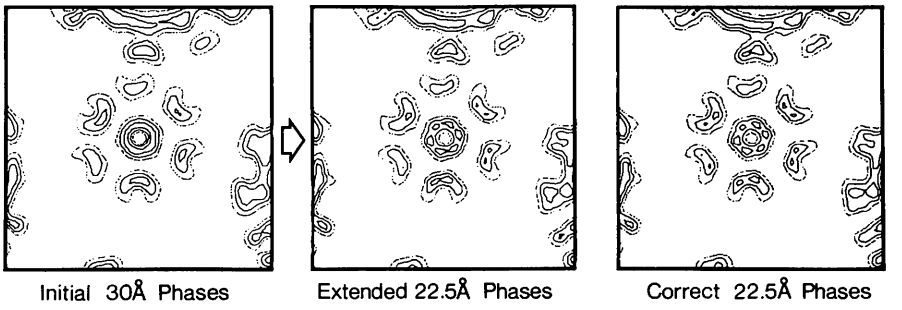


Fig. 8. Phase extension from 30 to 22·5 Å resolution.

the model amplitudes. This resultant data set was used in a series of phase refinements.

The error applied to the intensity for each reflection had two components. The first component was independent of the magnitude of the intensity and exhibited a Gaussian distribution about zero. This term mimicked the deviation in the film background which is a major term in the errors associated with the real data. The second component applied an error proportional to the magnitude of the intensities. The parameters were adjusted until the inaccurate data set gave a distribution of R factor with intensity and resolution that was worse than the scaling statistics observed for the real polyoma capsid data. The overall R factor between the inaccurate and perfect data was 10% on intensity which is worse than the symmetry-equivalent R factor for the capsid data of 5·7%.

The phases from the reasonable model were refined against the inaccurate data. The convergence was somewhat slower; however, after 15 cycles the overall R factor between the calculated and inaccurate set was 9·3%. This is approximately 1·5% higher than the convergence obtained with the native data. Most significantly, the r.m.s. phase difference between the refined and native model phases was only 31°. This is very similar to the value of 27° for the same refinement using the native data. These experiments show that small random errors do not prevent the refinement from converging close to the true phases.

Phase extension with the inaccurate structure factors converged to a final phase set whose final R factor was 8·8%, 2% higher than with the native model data. The r.m.s. phase difference between the refined and true phases was 35%, which is only slightly worse than the 32° r.m.s. phase difference observed for extension of the perfect data. The resultant electron density map still showed sixfold substructure in the hexavalent morphological units, even though there was a loss of some detail. In this case averaging the final electron density map enhanced the substructure of the morphological units.

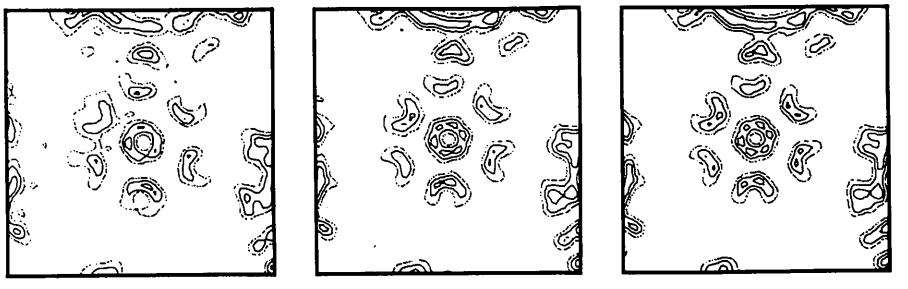
Weighting schemes

As noted by Bricogne (1976) it is desirable to weight each reflection in the map calculation according to the error in its phase. In this way reflections whose phases are correct will help refine those whose phase error is higher. For isomorphous replacement it is possible to derive a phase probability function and use this to weight the contribution of each reflection to the initial Fourier calculation. Such a function is not available for phases derived from model building; however, it was found that the simple weighting scheme described earlier greatly facilitated the phase refinement.

The primary purpose of the weighting scheme in the refinement of model phases is to allow the phases to change more easily and prevent the refinement from being locked into a false minimum. This freedom of movement in the phases is necessary because phases generated from models constitute a self-consistent set which impose a systematic bias on the initial map calculation. This is in contrast to phases generated by isomorphous replacement in which the errors should be randomly distributed about the correct set. Refinement of isomorphous phases primarily involves reducing the noise in the initial set.

The improvement in the convergence caused by the weighting scheme was demonstrated by refining the phases from the rudimentary model described earlier using unitary weights. Initially 42% of the low-resolution centric terms had incorrect signs. After 21 cycles of refinement with a unitary weighting scheme, the R factor between the calculated and native model amplitudes was stationary at 18% with an r.m.s. phase difference of 52°. This should be compared to the final R factor and r.m.s. phase difference of 7·2% and 28° when a weighting scheme was used. The major reason for the lack of convergence was the inability of a few large centric reflections with spacings between 80 and 100 Å to change signs to their correct values.

The unitary weighting scheme was useful once convergence using the exponential scheme had been



(*a*) (*c*) (*b*)

Fig. 10. The effect of unmeasured data on phase extension from 30 to 22·5 resolution. (*a*) Phase extension with 10% of the data missing and all terms with spacings >200 Å; (*b*) phase extension including calculated values for the missing data; (*c*) the correct native model map at 22·5 Å resolution.

achieved. A few additional cycles using unitary weights lowered the final R factor by ~0·5% and reduced the r.m.s. phase difference by ~4°. Thus, some form of weighting scheme dependent upon the fit of the calculated-to-observed amplitudes is necessary to allow the phases to change easily in the early stages of the refinement.

Solvent density values

The value given to the solvent density affects the calculated amplitudes and phases of the low-resolution structure factors. F_{000} was not included in any of the maps calculated for this study; thus, necessarily the total density in the maps was always zero. The solvent regions of the map were given their average value in order to maintain the zero total electron density. Since the envelope contains a net positive density the solvent value will be numerically negative. This was the simplest and most practical method of defining the solvent density.

The value given to the solvent density affected the phase refinement. Experiments with the native model data, in which the solvent density was set numerically to zero, failed to converge. This occurs because setting the density to zero, when numerically its average is negative, adds a net positive density to the map. The Fourier transform of the resultant map will contain a contribution in the calculated structure factors from this density. This will be the convolution of the transform of the solvent regions on the transform of the density. The effect will not be significant at high resolution; however, the very-low-resolution calculated terms will be incorrect. This has two consequences. Firstly, it will affect the scaling of the calculated to observed amplitudes and, secondly, the calculated values for unmeasured data will be incorrect.

Series termination results in diffraction ripples which extend throughout the map. This produces negative ripples below the mean solvent density. In those cases where the envelope contains significant regions of solvent (such as between the morphological units) negative ripples can occur. It is incorrect to set these points to the mean solvent density since this adds a net positive density to the interior of the envelope. This generates a spurious contribution to the calculated amplitudes. A better approach for accommodating these negative ripples is to create a more detailed envelope so that both the positive and negative components of the ripples can be removed.

Conclusion

The calculations that have been described show that trial phases from models of icosahedral particles such as polyoma virus (which are constructed using the structural information available from electron microscopy and small-angle scattering) can be refined successfully against low-resolution X-ray diffraction data using the constraints of solvent flattening and noncrystallographic symmetry averaging. The experiments show that a low R factor (~10%) between the calculated and observed amplitudes corresponds to a small phase difference between the refined and true phases, if the constraints are stringently applied. Thus, the R factor can be used to judge the progress of the refinement and assess the reliability of the phases on convergence. The calculations show that the initial phases do not continue to bias the refined phase set once a low R factor between the calculated and observed amplitudes has been achieved. From these studies it can be inferred that the refinement techniques are sensitive enough to distinguish between a hexameric and pentameric capsomer for the real polyoma virus capsid.

The model calculations also demonstrate the importance of compensating for any unmeasured data. They show that at low resolution it is essential to include calculated values for unrecorded data in the refinement calculations. Failure to reconstruct the unmeasured data will lead to a high R factor, phase error and confusing electron density maps.

The purpose of the calculations was to establish and verify the correct strategy for phasing the real polyoma virus capsid data. However, the observations on the treatment of unmeasured data, phase extension and the intrinsic limitations will apply in principle to any study that uses noncrystallographic symmetry and solvent flattening as a phase constraint. The use of model calculations which closely simulate an experimental system is a powerful technique for determining the reliability of the computational method and setting a level of expectation for the results.

I am indebted to Dr D. L. D. Caspar for insight, support and encouragement. I also thank Drs G. N. Phillips Jr, J. P. Fillers, and T. S. Baker for helpful discussions. The molecular replacement programs, together with many useful suggestions, were provided by Dr J. E. Johnson (Purdue University). The refinement calculations were performed at the Brandeis University Computer Center on a PDP-10. In addition, I thank Drs S. C. Harrison and I. K. Robinson (Harvard University) for demonstrating the numerical equivalence of the averaging programs written by G. Bricogne with those written by J. E. Johnson. This work was supported by a Young Investigators Research Grant CA27260, by PHS Grant CA15468 to D. L. D. Caspar from the National Cancer Institute and by NSF Grant PCM79-22766 for computer instrumentation.

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.

ABAD-ZAPATERO, C., ABDEL-MEGUID, S. S., JOHNSON, J. E., LESLIE, A. G. W., RAYMENT, I., ROSSMANN, M. G., SUCK, D. & TSUKIHARA, T. (1981). *Acta Cryst.* B**37**, 2002–2018.

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. (1974). *Acta Cryst.* A**30**, 395–405.

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

CASPAR, D. L. D. & KLUG, A. (1962). *Cold Spring Harbor Symp. Quant. Biol.* **27**, 1–24.

FINCH, J. T. (1974). *J. Gen. Virol.* **24**, 359–364.

HARRISON, S. C. (1971). *Cold Spring Harbor Symp. Quant. Biol.* **36**, 495–501.

HARRISON, S. C. & JACK, A. (1975). *J. Mol. Biol.* **97**, 173–191.

HARRISON, S. C., OLSON, A. J., SCHUTT, 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.

MAIN, P. (1967). *Acta Cryst.* **23**, 50–54.

NORDMAN, C. E. (1980). *Acta Cryst.* A**36**, 747–754.

RAYMENT, I., BAKER, T. S., CASPAR, D. L. & MURAKAMI, W. T. (1982). *Nature (London)*, **295**, 110–115.

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

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

UNGE, T., LILJAS, L., STRANDBERG, B., VAARA, I., KANNAN, K. K., FRIDBORG, K., NORDMAN, C. E. & LENTZ, P. J. JR (1980). *Nature (London)*, **285**, 373–377.

Acta Cryst. (1983). A**39**, 116–122

Effect of Dislocation Density on Integrated Intensity of X-ray Scattering by Silicon Crystals in Laue Geometry

BY N. M. OLEKHNOVICH, A. L. KARPEI, A. I. OLEKHNOVICH AND L. D. PUZENKOVA

Institute of Physics of Solids and Semiconductors, Byelorussian Academy of Sciences, Minsk 220726, *USSR*

(*Received* 29 *March* 1982; *accepted* 16 *August* 1982)

Abstract

An investigation of integrated intensity is performed for reflections 111 and 333 in plane-polarized Cu $K\alpha_1$ radiation for a series of silicon dislocation single crystals. Integrated intensity thickness oscillations (*Pendellösung* effect) have been found at low dislocation density (10–100 mm^{-2}). It is shown that the oscillations attenuate with increasing dislocation density, while their period somewhat increases. Thickness dependence of both extinction factor and polarization ratio is derived at high dislocation density (10^3–10^6 mm^{-2}). The present theoretical approaches based on the Darwin transfer equations appeared to be unsuitable for treating the obtained experimental data. They are analysed on the basis of coherent and diffuse scattering components.

1. Introduction

The previous investigations of polarization properties of Bragg reflections for silicon and germanium dislocation crystals (Olekhnovich, Markovich & Olekhnovich, 1980) show that the mosaic model of crystals is applicable for describing diffraction in real crystals, provided the dislocation density is over 10^4 mm^{-2}. Besides, diffraction in mosaic crystals is found to be practically determined only by primary extinction.

Kato (1980*a*,*b*), using equations of Takagi–Taupin type (Takagi, 1969), developed a statistical dynamical diffraction theory for crystals of any perfection degree. In this theory extinction is not subdivided into primary and secondary.

To establish the scattering mechanism of X-rays in real crystals it is important to study diffraction properties as a function of sample thickness using the Laue method. That method, as is known, allows one to investigate the *Pendellösung* effect, anomalous transmission, as well as the extinction effect. Lawrence & Mathieson (1977) proposed a simple method of single-crystal sample inclination for a controllable change of X-ray path length in Laue geometry. This procedure was used for studying integrated intensity thickness oscillation in perfect crystals (Somenkov, Shilstein, Belova & Utemisov, 1978).