Acta Cryst. (1985). A41, 252-262

Solving the Phase Problem in Fiber Diffraction. Application to Tobacco Mosaic Virus at 3.6 Å Resolution

BY KEIICHI NAMBA*

Rosenstiel Research Center, Brandeis University, Waltham, MA 02254, *USA*

AND GERALD STUBBS

Department of Molecular Biology, Vanderbilt University, Nashville, TN 37235, *USA*

(*Received* 10 *August* 1984; *accepted* 7 *November* 1984)

Abstract

The diffracting particles that give rise to a fiber diffraction pattern are randomly oriented about the fiber axis and, in consequence, the diffraction pattern is cylindrically averaged. The phase problem in fiber diffraction is not only to determine the phase in the usual crystallographic sense, but to overcome the loss of information from this averaging. This has been done by a multi-dimensional analog of protein crystallographic isomorphous replacement, combined with the use of information from the fine splitting of layer lines that occurs when a helical structure repeats approximately, but not exactly, in a given number of turns. The phases thus determined have been refined by a solvent-flattening procedure. They have been further refined by assuming the separation of cylindrically averaged Bessel-order terms (from multi-dimensional isomorphous replacement at an early stage and from a model at a later stage) and applying conventional isomorphous replacement (two-dimensional isomorphous replacement) to determine the phases of the terms. Cycles of model building and two-dimensional isomorphous replacement were found in the case of tobacco mosaic virus to improve greatly the quality of the electron density map, and enabled an atomic model of the virus to be built based on a highly interpretable map at 3·6 Å resolution with five Bessel orders (terms overlapping because of cylindrical averaging) separated.

Introduction

1. *The phase problem in fiber diffraction*

The phase problem in fiber diffraction is inherently more complicated than the phase problem in crystallography. In a crystal, the diffracting units all have the same orientation, but in a typical fiber diffraction experiment, although the diffracting particles are all oriented parallel to an axis, they are randomly oriented about that axis. For example, tobacco mosaic virus (TMV) particles, which are rod shaped, may be drawn into a capillary tube in a gel, 20-30% virus by weight, and by suitable manipulation (Gregory & Holmes, 1965) induced to line up with their long axes parallel to the capillary axis. However, they are randomly oriented about these axes.

The effect of this random orientation is to average the diffraction pattern cylindrically. The diffracted intensity at reciprocal-space radius R on layer line l, $I(R, l)$, is then

$$I(R, l) = \sum_n \mathbf{G}_{n,l}(R)\mathbf{G}^*_{n,l}(R) \qquad (1)$$

(Waser, 1955; Franklin & Klug, 1955). **G** is a complex Fourier-Bessel structure factor (Klug, Crick & Wyckoff, 1958), and the phase problem in fiber diffraction is to determine not only the phase of each **G**, but also the decomposition of I into $\{\mathbf{G}\}$.

2. *The number of terms to be separated*

For a given R, the sum in (1) contains only a finite number of significant terms, the number depending on the maximum radius and the symmetry of the diffracting system. If the diffracting structure is a helix with u identical subunits in t turns, where u and t are integers, the terms contributing to the sum are restricted to values of n and l that satisfy the selection rule

$$l = tn + um,$$

where m is integral (Cochran, Crick & Vand, 1952). Each $\mathbf{G}_{n,l}$ is a function of the Bessel function J_n:

$$\mathbf{G}_{n,l}(R) = \sum_j f_j J_n(2\pi R r_j) \exp i(-n\varphi_j + 2\pi l z_j/c). \qquad (2)$$

(r_j, φ_j and z_j are the cylindrical coordinates of atom j, which has a scattering factor f_j, and c is the length of the repeating unit of the structure in the z direction.) The finite radius of the diffracting structure and the fact that a Bessel function has an insignificant value until the argument approaches the order combine to restrict the number of terms in (1).

* Present address: Department of Molecular Biology, Vanderbilt University, Nashville, TN 37235, USA.

0108-7673/85/030252-11$01.50

For tobacco mosaic virus at 10 Å resolution, each layer line contains one or two significant **G** terms. At 3·6 Å, there are up to five, although only the outer parts of the structure contribute to the fifth term. For comparison, the filamentous bacteriophage pf1 has four overlapping terms in its diffraction pattern at 3·6 Å resolution (Makowski, Caspar & Marvin, 1980) whereas microtubules, with much lower symmetry, already have five overlapping terms at only 25 Å resolution (Beese, Stubbs & Cohen, 1985). It is possible to ignore the contributions of some of the higher-order terms, but there is a consequent loss of resolution, particularly in the outer parts of the structure.

3. *The phase problem for TMV*

Tobacco mosaic virus has served as the archetypal test system in the search for solutions to the fiber diffraction phase problem. At 10 Å it is possible to neglect the second **G** term and assume overlap is not present, and this enabled Barrett *et al.* (1971) to calculate an electron density map using isomorphous replacement as it is usually applied in protein crystallography. Holmes, Stubbs, Mandelkow & Gallwitz (1975) used an extension of isomorphous replacement into multiple dimensions (Stubbs & Diamond, 1975) to separate two Bessel orders (**G** terms) and calculate a 6·7 Å map. This work was extended to 4 Å and three Bessel orders by Stubbs, Warren & Holmes (1977).

Multi-dimensional isomorphous replacement is in many ways analogous to conventional protein crystallographic isomorphous replacement, but instead of two unknowns (the real and imaginary parts of the structure factor), there are $2n$ unknowns: the real and imaginary parts of each of n significant **G** terms contributing to the intensity. Consequently, large numbers of heavy-atom derivatives are required; at least two for each term to be separated. Stubbs & Makowski (1982) showed that the fine splitting of layer lines could be used to increase the amount of phase information available from each derivative, substantially increasing the resolution attainable with a limited number of derivatives. As an example, they calculated a 6·7 Å map of TMV using data from only two derivatives. (To calculate such a map by multi-dimensional isomorphous replacement alone would have required four derivatives.) In this paper, we present the methods, including multi-dimensional isomorphous replacement and layer-line splitting, that we have used in the calculation of a map of TMV at 3·6 Å resolution, considering contributions from up to five Bessel orders. This is not only an extension of the resolution of the 4 Å map; it is a major improvement over that map because it has not been necessary to neglect the contribution of any significant Bessel order.

Data collection

1. *Diffraction patterns*

Data were collected for TMV and the six heavy-atom derivatives described by Stubbs, Warren & Holmes (1977). Diffraction patterns were recorded photographically on Ilford Industrial G or Kodirex film in a Guinier camera using a point-focused beam from two bent quartz or germanium crystals and an Elliot rotating-anode generator (GX6 or GX13). Because of the convergence of the beam in this system, not more than two films were used in a pack, and a complete data set required data from at least two exposures. The films were scanned with an Optronics Photoscan densitometer, using a 50 μm raster.

2. *Determination of intensities*

The diffraction patterns obtained were of the type shown in Fig. 1. They consist of layer lines, and at high resolution these layer lines spread into each other. To obtain reliable estimates of intensities and layer-line positions from such patterns, the method of angular deconvolution (Makowski, 1978) was used, but modified in two ways. Firstly, the data were transformed from film space into reciprocal space (Fraser, Macrae, Miller & Rowlands, 1976) and corrected for non-linearity of film response and geometric factors before deconvolution. The assumption that all layer lines have the same angular cross

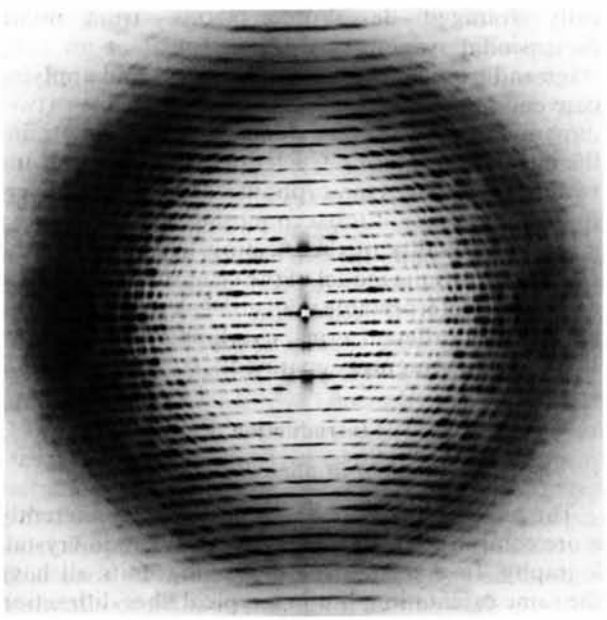

Fig. 1. Diffraction patterns from an oriented gel of tobacco mosaic virus, taken using flat film and a double-mirror focusing system. This unusually well oriented specimen was prepared in 1960 by K. C. Holmes and D. L. D. Caspar in Dr Caspar's laboratory. The pattern was recorded at the Rosentiel Center, Brandeis University, in 1982.

section (in most cases, a Gaussian) holds much better in reciprocal space. Secondly, in the rare cases where the predicted cross section was not accurate, such as the very near-meridional parts of layer lines 3 and 6, numerical integration was used. In order to monitor continuously the quality of the deconvolution and to make decisions about such factors as background shape and numerical integration, an interactive display was incorporated into the angular deconvolution program.

Determination of the background shape is an essential part of angular deconvolution. We generally used a background with circular symmetry. The transformation from Guinier film space to reciprocal space would produce a variable background from that part of the background attributable to the film fog, so the fog was estimated from an unexposed part of the film and subtracted before the transformation. X-rays scattered from slits occasionally produce a streak along the meridian, but this could be sufficiently well approximated by a Gaussian of appropriate width added to the constant background term. (Such an added term can only be used in angular deconvolution if its width is of a different order of magnitude from the layer-line width.) The possibility of using extra terms in the background, for example a Fourier series (Makowski, 1978) was considered, but with these terms present near-equatorial intensities were often underestimated (as judged from regions where the intensity of one or more layer lines was fortuitously low), and it appeared that the errors thus introduced were significantly larger than the very small errors the procedure was intended to correct.

Intensities were corrected for polarization effects after deconvolution.

3. *Scaling*

Films in a data set were scaled together after averaging the quadrants, using a Fourier surface whose coefficients (usually for five orders) were determined by least-squares methods (Weissman, Stauffacher & Eisenberg, 1985). R factors (on $I^{1/2}$) between the two strongest films were usually about 5%.

Derivative data sets were initially scaled to the native data set in the same way, with R factors varying from 10 to 20%. These data were then divided by 0.9, thus making some allowance for the heavy-atom scattering. Any local scaling procedure risks partial removal of the heavy-atom signal and, although this is considered in choosing the number of coefficients to describe the scaling surface, tests with simulated data sets showed that it was not possible to eliminate systematic errors in scaling due to the heavy atom contribution. To correct this problem, the data sets were rescaled after model building had begun (see below), using the current best estimate of the phases, and taking the calculated heavy-atom contribution into consideration.

Methods used in phase determination

This section will describe the methods used and referred to in later sections to determine preliminary phases, to improve these phases, and to refine heavy-atom parameters for use in phase determination. Note that where no confusion can arise, the word 'phase' is used to refer to the division of intensity into orthogonal parts, that is, the real and imaginary parts of each **G**.

1. *Phase determination*

(*a*) *Isomorphous replacement with layer-line splitting.* Equation (1) may be rewritten

$$I_j = \sum_{i=1}^{2N} (A_i + a_{j,i})^2. \qquad (3)$$

Here, I_j is the intensity at a given (R, l) for derivative j. N is the number of significant **G** terms contributing to the intensity at this (R, l), A_i is a real *or* imaginary part of $\mathbf{G}_{n,l}(R)$ for the native diffracted intensity, and $a_{j,i}$ is the heavy-atom contribution to that part of $\mathbf{G}_{n,l}(R)$ (zero in the case of the native). An independent equation of this type is available from each derivative, and from the native.

When the apparent repeat distance of the scattering particles (reflected in the layer-line spacing) corresponds only to an approximate repeat, the exact repeat distance being much longer (for example, in TMV there are actually 49·02 subunits in 3 turns, not 49), the Bessel function terms in each layer line do not fall at exactly the same values of Z in reciprocal space, that is, the layer line is split (Franklin & Klug, 1955). Because of disorientation and the finite thickness of layer lines, the splitting as such is not seen, but its effect is to cause very small shifts in the apparent position of the layer line, depending on the relative magnitudes of the contributing terms. These shifts are measured in the angular deconvolution, and used in the equation

$$\Phi_j I_j = \sum_{i=1}^{2N} q_j \varphi_i (A_i + A_{j,i})^2, \qquad (4)$$

where Φ_j is the angular shift observed for derivative j, φ_i is the calculated shift for the term corresponding to A_i, and q_j is the ratio of the degree of splitting in derivative j to that in the native (Stubbs & Makowski, 1982). An equation of this type is available from each data set for which specimen orientation is good enough to determine Φ_j reliably, and these equations are independent of (3).

The phase problem in this case is to determine $\{A_i\}$. We used linear equations derived from (3) and (4) to obtain a preliminary solution (Stubbs & Makowski, 1982), but we then used (3) and (4) directly to refine this solution. We made this change from the method of Stubbs & Makowski because of the difficulty of

assigning relative weights to the linear and non-linear equations, but the change was found to have little effect on the solution. We used a steepest-descents procedure to minimize the errors in (3) and (4). In this way, up to five Bessel-order terms were separated and phased, using intensities from six derivatives and splitting from four.

(*b*) *Improvement of the phase of each Bessel-order term.* After the steepest-descents procedure had been completed, a further phasing step was found empirically to improve the quality of the electron density maps, as judged by interpretability and chain continuity. Although it is not practical to search all of n-dimensional space ($n \leq 10$ in this case) to find the $\{A_i\}$ that minimizes the least-squares residual, a partial search was made by taking as fixed the values of $\mathbf{GG}^*$ (that is, the 'intensity' from each Bessel order) determined in the previous step, and determining the phase for each $\mathbf{G}$ term by the standard protein crystallographic procedure (see, for example, Blundell & Johnson, 1976, pp. 363–371). $\mathbf{G}$ was calculated for each derivative by taking the value of $\mathbf{G}$ for the native as determined in the previous step, and adding the heavy-atom contribution. This calculated $\{\mathbf{G}\}$ was normalized so that $\sum \mathbf{GG}^*$ was equal to the observed derivative intensity. Since this two-dimensional problem is significantly over-determined by the intensity data from six derivatives, layer-line splitting data were not used in this step. Phases determined in this way could be most probable phases (which were used during the refinement of heavy-atom parameters) or best phases (used for calculating maps). Figures of merit were also available, and although it must be pointed out that these figures of merit do not have exactly the same meaning as those in crystallography, since they take no direct account of errors in the decomposition of I into $\{\mathbf{G}\}$, they do reflect the quality of the phase determination at each point in the diffraction pattern. This procedure will be referred to below as 'two-dimensional isomorphous replacement'.

It is not completely clear why this procedure should improve the maps, but the most likely possibilities are that the steepest-descents procedure may lead to false or incompletely determined minima, and that it is difficult to weight correctly the effect of layer-line splitting on the phase. For example, at high values of R near the equator, splitting data are less reliable than intensity data, and two-dimensional isomorphous replacement using intensity data alone somewhat reduces the dependency of the phase on these data.

2. *Phase refinement*

The phases determined by isomorphous replacement and layer-line splitting inevitably contain errors due to imperfect heavy-atom positions, scaling and other factors. Extra information about the structure can be used to constrain and thus improve these phases. In protein crystallography, real-space averaging of the electron density of molecules related by non-crystallographic symmetry and flattening of the electron density in solvent regions have been shown to improve phases greatly in suitable cases (for example, Bloomer, Champness, Bricogne, Staden & Klug 1978; Harrison, Olson, Schutt, Winkler & Bricogne 1978). Two related methods of phase refinement were used in this work, both analogous to the protein crystallographic solvent-flattening procedure (as used, for example, by Wang, 1985; Phillips, Lattman, Cummins, Lee & Cohen 1979).

(*a*) *Box-function refinement.* In this procedure (Makowski, 1981, 1982), the electron density map [or, in practice, the Fourier–Bessel components $g_{n,l}(r)$ of the map] is truncated, that is, multiplied by a box, at the known outer radius of the particle. It is then Fourier–Bessel transformed to give $\{\mathbf{G}\}$, the decomposition of I and the phases of this $\{\mathbf{G}\}$ are applied to the observed intensities, and a back transformation gives a new model. Two or three cycles of this process usually suffice for convergence. This method of refinement is equivalent to imposing the constraint that the orthogonal parts $\{A_i\}$ be continuous, that is, that the frequency of these fluctuations is bounded, in accordance with the 'minimum wavelength' principle (Bragg & Perutz, 1952).

(*b*) *Solvent flattening.* Box-function phase refinement is a solvent-flattening procedure, which has been useful in fiber diffraction, but since the cylindrical box is a very rough definition of the molecular boundary, the constraint on the phase is weaker than a more complete solvent flattening would be. Furthermore, in the box-function refinement each layer line is refined independently, so if one layer line refines to a wrong phase set, no amount of improvement of the other layer lines will correct it. We therefore used a full solvent-flattening procedure at this point.

Solvent flattening requires three steps: definition of a molecular boundary, modification of electron density outside (and sometimes inside) this boundary, and refinement of the phases on the basis of this modified density. Our procedure for defining the molecular boundary is basically the same as that of Wang (1985). Electron densities lower than some selected (usually small positive) contour value are set to zero, and densities are locally averaged to make a lower-resolution representation of the molecular density. This local averaging can be done in various ways; our procedure was to transform the map to reciprocal space and then back transform it using an artificial temperature factor of 300 Å^2. The molecular boundary is then defined as some contour, still positive but lower than the first one used. The values for

initial contouring and for drawing the boundary are subjectively judged by comparing the boundary with the original map.

Electron density outside the boundary is set to zero, and any deep holes, such as sometimes occur at heavy-atom sites, are truncated to some moderate depth. The refinement then proceeds analogously to box-function refinement: the modified map is transformed, the calculated phases are assigned to the observed diffracted intensities and the electron density is again calculated. Two cycles of the procedure (without recalculating the boundary) are usually enough to obtain convergence.

Although this refinement is similar to box-function refinement, it requires more than three times as much computer time, so it was not used in the heavy-atom parameter refinement described below. However, it imposes a considerably stronger constraint on the phases, and inspection of the maps thus produced suggested that its use at a later stage improved the phases significantly.

(*c*) *Combination of phasing methods. Sim weighting.* In the last step of the phase determination (see below) the phase probabilities from two-dimensional isomorphous replacement were combined with probabilities based on the solvent-flattening refined map, using a weighting analogous to that described by Sim (1959). The probability of the phase α falling in the range $\alpha, \alpha+\mathrm{d}\alpha$ is taken as

$$P(\alpha)\mathrm{d}\alpha = N \exp\left[\frac{2G_oG_c}{\Sigma}\cos(\alpha-\alpha_c)\right]\mathrm{d}\alpha,$$

where G_o is the currently accepted value of G, G_c is calculated from the refined map, α_c is the phase of $\mathbf{G}_c$, Σ is an estimate of the mean-square error in $\mathbf{G}_c$ and N is a normalizing factor (Stubbs, 1972; Blundell & Johnson, 1976, p. 419). Σ can be estimated as $\langle G_o^2 - G_c^2\rangle$, averaged within a range of resolution, but, because of the multi-dimensional nature of this problem, this is a significantly low estimate. We therefore used a factor c so that $\Sigma = c\langle G_o^2 - G_c^2\rangle$. A satisfactory value of c (judged from the quality of the resulting maps) was found to be 20.

3. *Refinement of heavy-atom parameters*

The heavy-atom parameters used at the beginning of this work were those determined by Stubbs, Warren & Holmes (1977). These parameters were refined using the least-squares method described for crystallography by Dickerson, Weinzierl & Palmer (1968). The relationship between the parameters and $\{\mathbf{G}\}$ is not identical to that between crystallographic parameters and $\{\mathbf{F}\}$, but we have shown by model calculations analogous to those of Dickerson, Weinzierl & Palmer that the algorithm works for cylindrically averaged data, provided that the starting phases are good enough.

Table 1. *Heavy-atom parameters*

For further details of the derivatives, see Stubbs, Warren & Holmes (1977).

Ligand	Strain	f	r	φ	z	B
MMN	Vulgare	88·4	57·01	15·30	12·49	27·0
MMN	Ni-2068	96·6	56·96	15·27	12·42	37·2
		73·0	72·23	0·37	1·07	76·0
Pb^{2+}	Vulgare	29·5	24·75	6·54	20·62	57·4
SHIMS-MNN	Vulgare	13·7	70·30	5·50	−1·29	50·0
		95·3	57·02	15·58	12·55	37·3
UO_2F_5	Vulgare	96·2	57·20	15·24	12·44	35·2
		41·7	26·64	7·14	19·72	86·9
		59·1	57·91	22·46	25·98	107·9
OsO_4	Vulgare	25·9	91·95	4·81	7·76	45·1

One cycle of this algorithm consists of two steps: phase determination and least-squares-parameter refinement. The initial phases were calculated by multi-dimensional isomorphous replacement with layer-line splitting, and the most probable phase for each Bessel-order term was derived by two-dimensional isomorphous replacement with the separation of the terms calculated from this initial $\{A_i\}$. At this stage, because of errors, $A_i(R)$ had discontinuities. Box-function refinement was applied to remove these discontinuities, and two-dimensional phasing was repeated. If the discontinuities persist, the box-function refinement should be repeated, but, in practice, $\{A_i(R)\}$ contained no major discontinuities by this point. Least-squares refinement of heavy-atom parameters was then based on this set. Three cycles of least-squares calculation were generally necessary for convergence.

Data between 20 and 4 Å resolution were used in the heavy-atom-parameter refinement, but data points whose residuals were in the highest 10% for any cycle were omitted from that cycle. Atomic positions and occupancies were refined. Temperature factors showed a tendency to take on unreasonably low values if they were allowed to vary freely, especially in the early stages of the work. In view of the complicated nature of the scaling surface, the scale coefficients were not refined. Instead, they were recalculated at intervals.

An attempt was made to refine the parameters before the start of the phasing process. Although most of the parameters did refine to stable values, the electron density maps were not improved, and scaling of the data sets considering the heavy-atom contribution served only to make them worse. In contrast, after the phasing process was complete and a model had been built, heavy-atom-parameter refinement based on two-dimensional isomorphous replacement with decomposition of I from the model worked very well (see below). At that point, scaling considering heavy atoms also began to improve the maps. It is clear that, in this system, heavy-atom-parameter refinement works, but only after sufficiently good starting phases have been obtained.

The refined heavy-atom parameters are given in Table 1.

Calculation of a 3·6 Å resolution map

Using the methods described above, a map was calculated from data extending to 3.6 Å resolution. Figs. 2(*b*)-(*e*) show a part of this map at various stages of refinement. For comparison, the 4 Å map (Stubbs, Warren & Holmes, 1977) is shown in Fig. 2(*a*).

(*a*) The first map (Fig. 2*b*) was calculated by multi-dimensional isomorphous replacement with layer-line splitting as described above. This map is already markedly superior to the 4 Å map, partly because of the increased resolution, but primarily because five Bessel orders have been used, rather than three. The region of the map shown in Fig. 2 is near the outside of the TMV particle, and is therefore particularly sensitive to the omission of high-order Bessel terms.

(*b*) A second map (Fig. 2*c*) was calculated by two-dimensional isomorphous replacement, taking the Bessel-order separations from (*a*). The map was improved in this step in several critical ways; in particular, the connectivity along the peptide chain was better, and therefore the interpretation was easier.

(*c*) The phases in (*b*) showed discontinuities along the layer lines, so the box-function refinement procedure was applied.

(*d*) Two-dimensional isomorphous replacement phasing was used as in (*b*), taking Bessel-order separations from (*c*), to calculate the map shown in Fig.

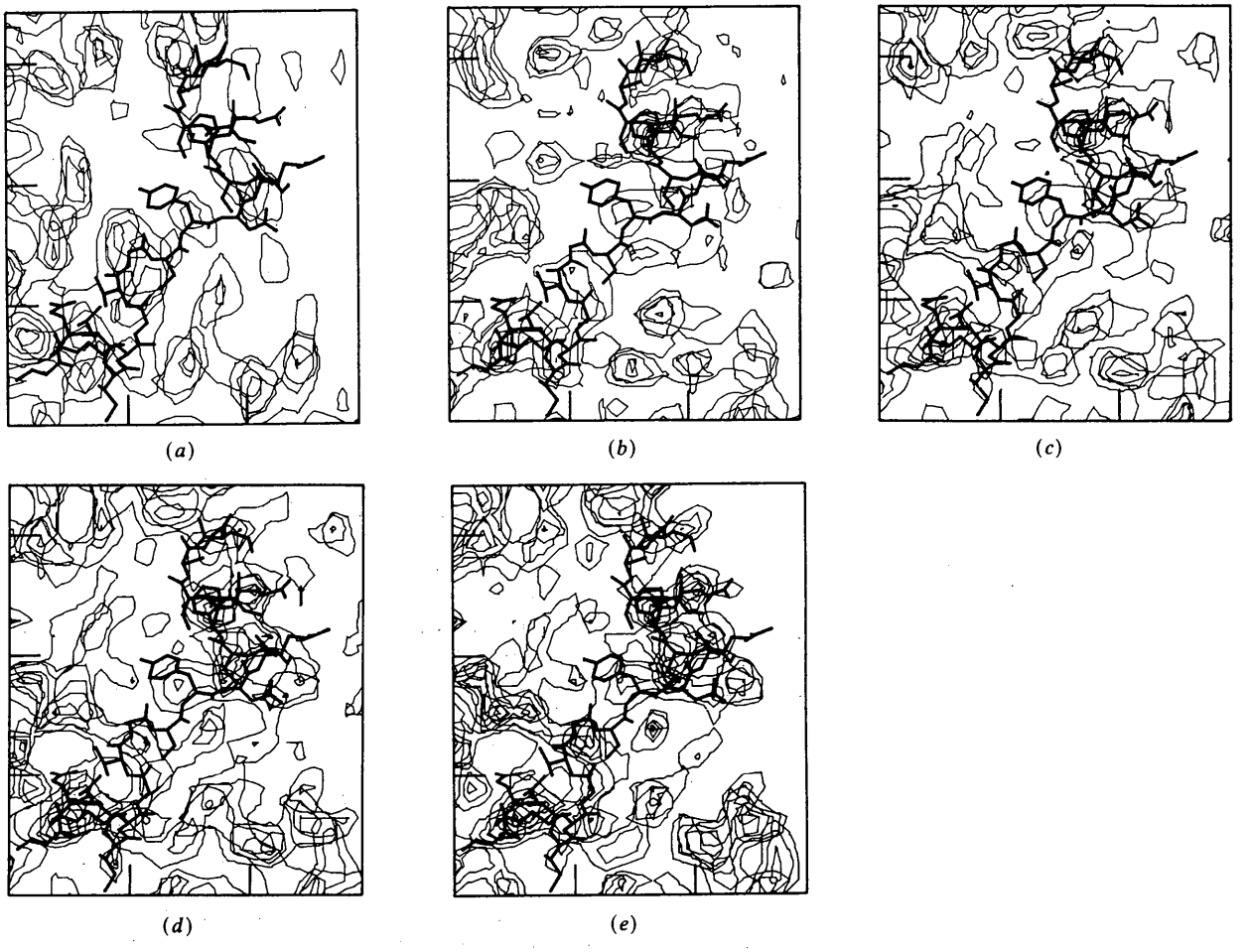

Fig. 2. Progressive improvement of the electron density map at various early stages of phase determination. The map shows an axial view of the outside part of the virus particle. A 4·2 Å thick slice is shown with an approximate model (not the final refined model) of residues 130 to 150 superimposed to indicate the course of the protein chain. Note the improvement in chain continuity and the increasing clarity of the turns of α-helix in the top right corner of the map. (*a*) The map calculated in 1977 (Stubbs *et al.*, 1977) at 4 Å resolution by multi-dimensional isomorphous replacement. Intensity data from the native virus and six heavy-atom derivatives were used. Contributions from up to three overlapping Bessel orders were considered. (*b*) The map at 3·6 Å resolution with contributions from five Bessel orders considered. Most of the improvement is due to the complete set of Bessel orders rather than the slight extension of resolution. Separation of the extra terms was made possible by considering the very fine splitting of the layer lines for the native and four of the six derivatives. (*c*) The map calculated by two-dimensional isomorphous replacement, taking the Bessel-order separations from (*b*) and using intensity data from the native and six heavy-atom derivatives. 'Best' phases were used. (*d*) The map after applying the box-function refinement to (*c*) and calculating two-dimensional isomorphous replacement phases. (*e*) The map after applying solvent-flattening refinement to (*d*) and calculating two-dimensional isomorphous replacement phases.

2(*d*). The mean phase change* between (*b*) and (*d*) was 29°.

(*e*) Two cycles of solvent-flattening refinement were applied to the map from (*d*).

(*f*) The final map in this series (Fig. 2*e*) was calculated by two-dimensional isomorphous replacement, taking the Bessel-order separation from (*e*). This time, however, the phases from solvent flattening were combined with the isomorphous replacement phases using the Sim weighting scheme described above. The mean phase change between (*d*) and (*f*) was 16°.

The sequence of procedures used to calculate this map was determined largely by experiment and examination of the successive maps. As can be seen in Fig. 2, the quality of the map was improved at each step, with a large improvement overall. This improvement is also indicated in Fig. 3, where figures of merit from the two-dimensional isomorphous-replacement calculations are shown for different resolution ranges. The greatest improvement in the figures of merit came with the solvent-flattening refinement, demonstrating that this procedure invokes a substantially stronger constraint than does the box-function refinement.

Model building

1. *Initial model building*

A molecular model was built to fit the map partially shown in Fig. 2(*e*), using the Evans and Sutherland graphics system (models PS2 and MPS) and the program *FRODO* (Jones, 1982). Such models have been built before for the inner part of the protein in the intact virus (Stubbs, Warren & Holmes, 1977), for the RNA (Stubbs & Stauffacher, 1981), and for the outer part of the protein in the disk aggregate of the protein (Bloomer *et al.*, 1978). All of these models were used as guides. Unlike the virus, the disk protein is disordered near the inner surface of the virus (residues 89–113). The main chain of the protein in the new model was extremely similar to that of the disk protein except in and near this region. The RNA model was only slightly changed from that of Stubbs & Stauffacher (1981).

The low-radius loop (residues 90–110) was rather difficult to build at this stage, and required considerable modification later in the refinement process, as the electron density map was initially rather noisy in this region. Nonetheless, it was possible to build an initial approximation to the loop, taking into account various constraints. These included charge interactions such as salt bridges, and the positions of the heavy atoms in the lead and uranyl fluoride derivatives. These atoms are believed to bind preferentially to carboxyl groups (Caspar, 1963; Blundell & Johnson, 1976). Use was also made of a derivative of the methionine mutant Ni 630 (Wittmann, 1962) with platinum; the occupancy was too low to permit the use of this derivative in the phasing, but a difference Fourier map contained a clear single peak, which was tentatively interpreted as a platinum ion binding to Met 107.

The crystallographic *R* factor for the initial model was 0·314 for data between 50 and 3·6 Å resolution.

2. *Refinement of the model*

Once the initial model had been built, structure factors were calculated from the atomic coordinates. The separation of the Bessel-order terms was taken from these structure factors and used to determine phases by two-dimensional isomorphous replacement, and the map from these phases was used to rebuild the model. This procedure was repeated five times, by which time the model was not changing significantly. After the first two such cycles, the heavy-atom parameters were refined again, the refinement being based on the most probable two-dimensional isomorphous replacement phase with Bessel-order separation from the model.

During these five cycles of phase and model refinement, the *R* factor fell to 0·287. More significantly, the map improved greatly. This is particularly well illustrated in Fig. 4, where we see the map and model at different stages of the refinement for a region including Ser 1, Trp 152, Thr 153 and Ser 154, on the outside surface of the particle. The improvement in

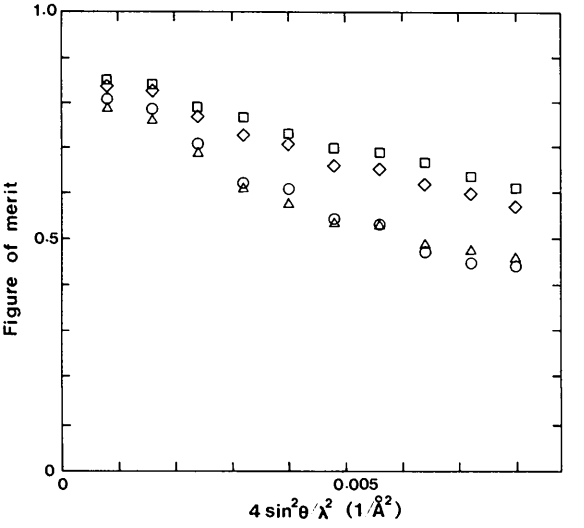

Fig. 3. Figures of merit at various stages of phase determination. ○ Phases corresponding to Fig. 2(*c*). △ Phases corresponding to Fig. 2(*d*). ◇ Two-dimensional isomorphous replacement phases after the solvent-flattening refinement had been applied to the map in Fig. 2(*d*). □ As ◇, but with phase probabilities combined with phase probabilities from the solvent-flattening refinement. These phases correspond to Fig. 2(*e*). Overall figures of merit are ○ 0·56, △ 0·56, ◇ 0·68, □ 0·71.

* Mean phase change is defined in this case as the angle between the 2*N*-dimensional vectors representing two different values of **G**.

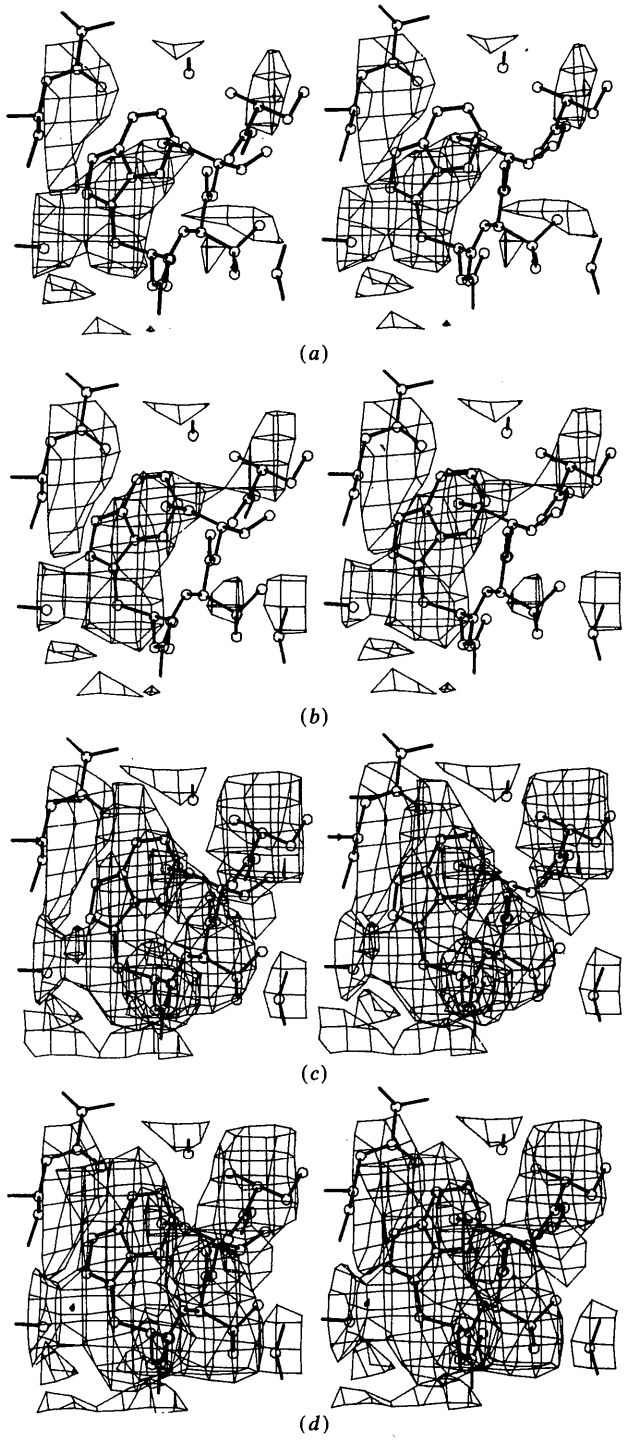

Fig. 4. Progressive improvement of the electron density map as the model and the heavy-atom parameters were refined. The maps were calculated from the two-dimensional isomorphous replacement best phase, taking the Bessel-order separations from the current model. The figure shows the final conformations of some of the residues near the N and C termini, at the outside surface of the virus. (*a*) After one cycle of model building and phase calculation. (*b*) After two cycles of model building and phase calculation. (*c*) After the second cycle, the heavy-atom parameters were refined and the phases were re-calculated. (*d*) After three cycles of model building and phase calculation.

the electron density derives solely from rebuilding other parts of the model, since the density in this region was initially so poor that the model could not be modified to fit it.

Although 3·6 Å is not usually considered to be sufficient resolution for the building of an atomic model, the examples shown in Fig. 5 show that after this extensive refinement process this map was indeed good enough to permit considerable confidence in the model. Fig. 5(*a*) shows an α-helical stretch of backbone, while Figs. 5(*b*) and (*c*) show side-chain interactions.

3. *Tests for bias in the map*

Since our refinement procedure depends on successive improvements in the model being incorporated into the Bessel-order separation (although not directly into the phasing of each term), it was of paramount importance to ensure that the apparent high quality of the map was not merely caused by bias towards the model. This is shown to some degree in Fig. 4, where the electron density near the N and C termini is shown at different stages of the refinement. The poor quality of the map in Fig. 4(*a*) shows that, at least in the initial phase calculation by two-dimensional isomorphous replacement, the influence of the unrefined model was not sufficient to produce an electron density map reflecting only the input model structure. This electron density improves considerably during the refinement, even though the corresponding part of the model could not be adjusted significantly because of the poor quality of the density early in the refinement. The independence of the map is shown much more directly in Fig. 6, which illustrates the effect of the model of the side chain of Tyr 70 on the electron density. The map calculated as described above, using two-dimensional isomorphous replacement with Bessel-order separation from the entire model shows density corresponding to the side chain (Fig. 6*a*). A map calculated directly from the model without the side chain has, as expected, no density around the side chain (not shown). A difference Fourier map (Mandelkow, Stubbs & Warren, 1981) shows a peak of density corresponding to the missing side chain (Fig. 6*b*). A map calculated with Bessel function separations from the partial model and with phases for each separate term from two-dimensional isomorphous replacement (Fig. 6*c*) agrees excellently with the map calculated from the full model, demonstrating that the phase information does not merely correspond to that contained in the current model.

Discussion

In this paper, we have repeatedly referred to the quality and interpretability of the map as the major criterion for assessing the value or otherwise of any

step in the refinement. The procedure has been somewhat empirical so far; further experience with this and other structures will be needed before we can expect to establish the optimal methods for determining a relatively large macromolecular structure at high resolution using fiber diffraction data. Nonetheless, we have obtained a highly interpretable map, and shown that it is not unduly biased towards the models used in its production.

For several reasons, this map and model are substantially improved from those described in 1977 (Stubbs, Warren & Holmes), although the nominal resolution has improved only from 4 to 3·6 Å. The data are better, particularly at high resolution, because the angular deconvolution procedure enabled us to make better estimates of intensities and backgrounds in the regions of substantial overlap of layer lines. The map is improved, particularly near the outer surface of the particle, because all five contributing Bessel terms were able to be considered, instead of only three. This was possible because of the use of layer-line splitting. Finally, the map was further improved by various refinement procedures, including two-dimensional isomorphous replacement, solvent flattening and the simultaneous refinement of the phases and the model described above.

The methods described in this paper have many potential applications in fiber diffraction. Very few systems include such well-oriented specimens and such detailed diffraction patterns as does TMV, but there are many that could benefit to some degree by this approach. Work is currently in progress in this laboratory on several strains of TMV, some of which orient as well as TMV itself. While the structures of such strains could be solved by difference Fourier-Bessel syntheses (Mandelkow, Stubbs & Warren, 1981), such difference syntheses are more biased towards the known structure in fiber diffraction than in crystallography. If one or two heavy-atom derivatives are available, two-dimensional isomorphous replacement can be used, taking the Bessel-order separation from the known structure. Our experience as described in this paper suggests that such a synthesis will be much less biased towards the TMV structure.

The model to be used in separating the Bessel orders need not be derived from multi-dimensional isomorphous replacement, either directly or as a related structure. It could be a model derived by refinement of a trial structure, as in the case of filamentous bacteriophage pf1 (Makowski, Caspar & Marvin, 1980), or even, at low resolution, simply a refined electron density map, as in the case of microtubules (Beese, Stubbs & Cohen, 1985). Thus the methods described here bring structures that have seemed intractable to isomorphous-replacement methods because of the difficulty of making derivatives, especially in large numbers, within reach.

Little has been said in this paper about the actual structure we have determined for TMV. Details of

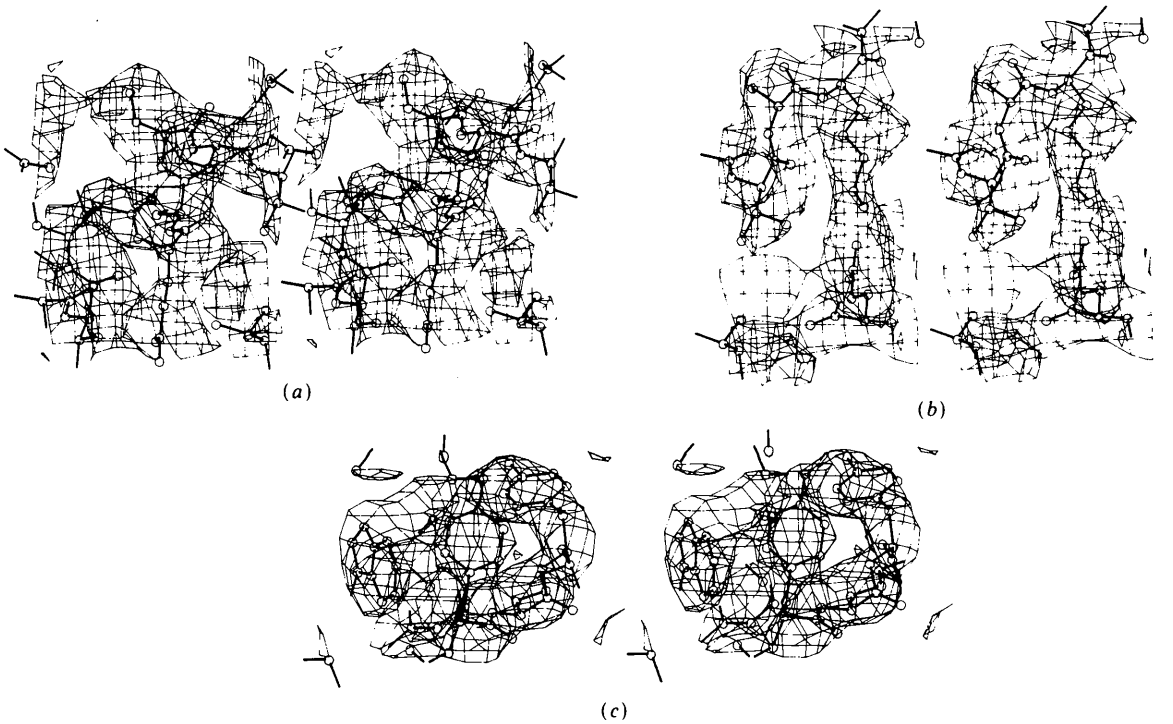

Fig. 5. Parts of the final electron density map with the final model superimposed. (*a*) Part of the right slewed α-helix (terminology according to Champness, Bloomer, Bricogne, Butler & Klug, 1976). (*b*) A salt bridge between Lys 53 and Glu 22. (*c*) Trp 17, Tyr 70 and Tyr 72 in close proximity.

this will be published elsewhere, but a few observations are in order. As already observed (Stubbs, Warren & Holmes, 1977; Bloomer *et al.*, 1978), the peptide backbones of the virus structure and the disk structure are extremely similar except in the vicinity of the RNA and inside the RNA radius. With more details visible, we find that the mean difference in position for the main-chain atoms for the rest of the structure is probably less than 0·5 Å. However, there are many substantial side-chain differences. Among the largest are those in the interface between successive turns of the virus helix or layers of the disk. This interface is quite different even at a gross structural level, since the subunits move by about half the width of one subunit during the disk-to-helix transition. In the disk, this interface includes a complex hydrogen-bonding network (Bloomer *et al.*, 1978). In the virus, a hydrogen-bonding network has also been predicted (Stubbs, Warren & Mandelkow, 1979). This network has now been shown to exist, but even at lower radius, where side-chain conformational changes might have been expected to be able to accommodate the change in quaternary structure, the specific hydrogen bonds appear (subject to the limitations of interpretation of a 3·6 Å map) to be quite different.

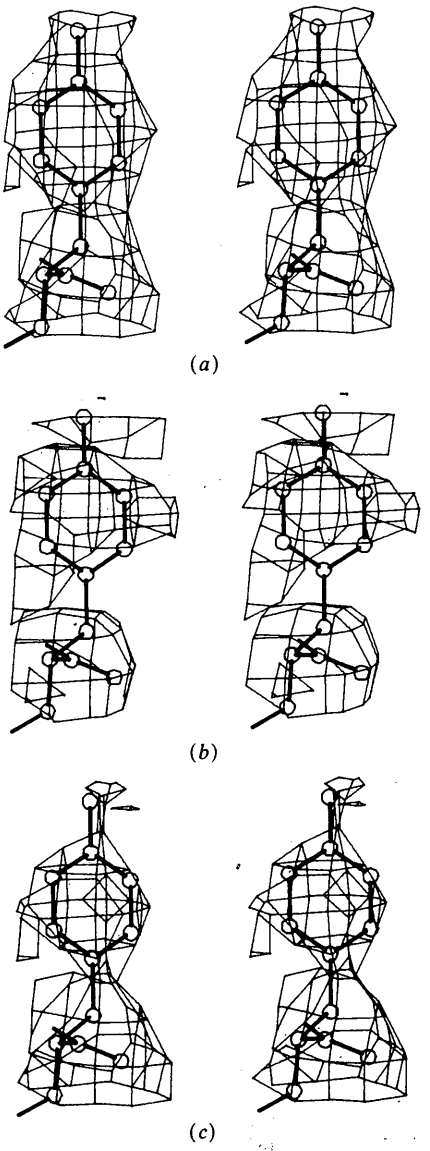

Fig. 6. Demonstration that the electron density map is not unduly biased towards the model. (*a*) A part of the final map showing Tyr 70. The map is calculated by two-dimensional isomorphous replacement, taking the Bessel-order separations from the entire model. (*b*) A difference Fourier map based on structure factors from the model without the side chain of Tyr 70. (*c*) A two-dimensional isomorphous replacement map, taking the Bessel-order separations from the model without the side chain of Tyr 70.

This work was supported by NIH grants GM 25236 and GM 33265. Funds to purchase and maintain the VAX 11/780 computer at Brandeis were from a NIH Shared Instrumentation Grant 3-R01-GM21189-09S1 awarded to D. J. DeRosier. We thank the European Molecular Biology Laboratory and the Department of Biochemistry, Columbia University, for the use of computer graphics facilities, and Dr Lee Makowski for critically reading the manuscript.

References

BARRET, A. N., BARRINGTON LEIGH, J., HOLMES, K. C., LEBERMAN, R., MANDELKOW, E., VON SENGBUSCH, P. & KLUG, A. (1971). *Cold Spring Harbor Symp. Quant. Biol.* **36**, 433-448.

BEESE, L., STUBBS, G. & COHEN, C. (1985). In preparation.

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

BLUNDELL, T. L. & JOHNSON, L. N. (1976). *Protein Crystallography.* New York: Academic Press.

BRAGG, W. L. & PERUTZ, M. F. (1952). *Proc. R. Soc. London Ser. A*, **213**, 425-435.

CASPAR, D. L. D. (1963). *Adv. Protein Chem.* **18**, 37-121.

CHAMPNESS, J. N., BLOOMER, A. C., BRICOGNE, C., BUTLER, P. J. G. & KLUG, A. (1976). *Nature (London)*, **259**, 20-24.

COCHRAN, W., CRICK, F. H. C. & VAND, V. (1952). *Acta Cryst.* **5**, 581-586.

DICKERSON, R. E., WEINZIERL, J. E. & PALMER, R. A. (1968). *Acta Cryst.* B**24**, 997-1003.

FRANKLIN, R. E. & KLUG, A. (1955). *Acta Cryst.* **8**, 777-780.

FRASER, R. D. B., MACRAE, T. P., MILLER, A. & ROWLANDS, R. J. (1976). *J. Appl. Cryst.* **9**, 81-94.

GREGORY, J. & HOLMES, K. C. (1965). *J. Mol. Biol.* **13**, 796-801.

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

HOLMES, K. C., STUBBS, G. J., MANDELKOW, E. & GALLWITZ, U. (1975). *Nature (London)*, **254**, 192-196.

JONES, T. A. (1982). *Computational Crystallography*, edited by D. SAYRE, pp. 303-317. Oxford Univ. Press.

KLUG, A., CRICK, F. H. C. & WYCKOFF, H. W. (1958). *Acta Cryst.* **11**, 199-213.

MAKOWSKI, L. (1978). *J. Appl. Cryst.* **11**, 273-283.

MAKOWSKI, L. (1981). *J. Appl. Cryst.* **14**, 160-168.

MAKOWSKI, L. (1982). *J. Appl. Cryst.* **15**, 546-557.

MAKOWSKI, L., CASPAR, D. L. D. & MARVIN, D. A. (1980). *J. Mol. Biol.* **140**, 149-181.

MANDELKOW, E., STUBBS, G. & WARREN, S. (1981). *J. Mol. Biol.* **152**, 375-386.

PHILLIPS, G. N., LATTMAN, E. E., CUMMINS, P., LEE, K. Y. & COHEN, C. (1979). *Nature (London)*, **278**, 413-417.
SIM, G. A. (1959). *Acta Cryst.* **12**, 813-815.
STUBBS, G. (1972). *Structural Studies of Crystalline Proteins.* Thesis, Univ. of Oxford.
STUBBS, G. & DIAMOND, R. (1975). *Acta Cryst.* A**31**, 709-718.
STUBBS, G. & MAKOWSKI, L. (1982). *Acta Cryst.* A**38**, 417-425.
STUBBS, G. & STAUFFACHER, C. (1981). *J. Mol. Biol.* **152**, 387-396.
STUBBS, G., WARREN, S. & HOLMES, K. (1977). *Nature (London)*, **267**, 216-221.
STUBBS, G., WARREN, S. & MANDELKOW, E. (1979). *J. Supramol. Struct.* **12**, 177-183.
WANG, B. C. (1985). In preparation.
WASER, J. (1955). *Acta Cryst.* **8**, 142-150.
WEISSMAN, L., STAUFFACHER, C. & EISENBERG, D. (1985). In preparation.
WITTMANN, H. G. (1962). *Z. Vererbungsl.* **93**, 491-530.

Acta Cryst. (1985). A**41**, 262-268

Search for a Fragment of Known Geometry by Integrated Patterson and Direct Methods

BY ERNST EGERT AND GEORGE M. SHELDRICK

Institut für Anorganische Chemie der Universität Göttingen, Tammannstrasse 4, D-3400 Göttingen, Federal Republic of Germany

(*Received 6 July* 1984; *accepted 26 November* 1984)

Dedicated to Professor Jack D. Dunitz on the occasion of his sixtieth birthday

Abstract

A method is presented that attempts to exploit all the *a priori* available information in order to locate a fragment of known geometry in the unit cell. Whereas the orientation of the search model is determined by a conventional but highly automated real-space Patterson rotation search, its position in the cell is found by maximizing the weighted sum of the cosines of a small number of strong translation-sensitive triple-phase invariants, starting from random positions. A Patterson minimum function based on intermolecular vectors is calculated only for those solutions that do not give rise to intermolecular contacts shorter than a preset minimum. This procedure avoids the time-consuming refinement in Patterson space and should be especially efficient for large structures. Finally, the best solutions are sorted according to a figure of merit based upon the agreement with the Patterson function, the triple-phase consistency and an R index involving E_{obs} and E_{calc}. Tests with about 30 known structures, using search fragments taken from other published structures or from force-field calculations, have indicated that this novel combination of Patterson and direct methods is reliable and widely applicable. A few selected examples demonstrate the power of the computer program *PATSEE*, which is compatible with *SHELX*84 and will be distributed together with it. *PATSEE* is valid and efficient for all space groups and imposes no limits on the number of atoms or data. The orientation search for a single fragment allows one additional degree of torsional freedom, and up to two fragments may be translated simultaneously.

Introduction

The choice of strategy for the solution of a crystal structure at atomic resolution is usually determined by the presence or absence of heavy atoms. Thus it is common practice to solve light-atom structures with direct methods and those containing heavy atoms with Patterson techniques. If this (very often straightforward) strategy fails, it may be advisable to resort to the corresponding alternative method: direct methods may well reveal the positions of heavy atoms, and the Patterson function can be interpreted even for purely light-atom structures, such as those of organic molecules, provided that part of the molecular geometry is known. This so-called Patterson search has been shown by various authors to be a powerful tool for solving difficult crystal structures; its great strength is that it employs chemical information directly, and so can compensate for mediocre precision and resolution of the X-ray data (Egert, 1983, and references cited therein). Nevertheless, it is not nearly as popular as direct methods, which owe part of their success to automation and superior computational efficiency. In this paper we describe an attempt to combine the merits of both methods - in a manner that is generally applicable, efficient, automatic, computer independent and easy to use - and thus to exploit *all* the *a priori* available information in order to solve large problem structures.

Preparation of the search

There are a number of different methods of performing a Patterson search, but they all fall into one of

0108-7673/85/030262-07$01.50