Acta Cryst. (1975). A31, 709

The Phase Problem for Cylindrically Averaged Diffraction Patterns. Solution by Isomorphous Replacement and Application to Tobacco Mosaic Virus

BY G. J. STUBBS

Max-Planck-Institut für medizinische Forschung, Heidelberg, Jahnstrasse 29, Germany (BRD)

WITH AN APPENDIX BY R. DIAMOND

Medical Research Council – Laboratory of Molecular Biology, Hills Road, Cambridge, England

AND G. J. STUBBS

(*Received* 11 *February* 1975; *accepted* 20 *May* 1975)

A procedure is given for separating and phasing terms originating from different Bessel-function orders in cylindrically averaged fibre diffraction patterns, using isomorphous heavy-atom derivatives. Model calculations and an application to tobacco mosaic virus are described.

In structural studies on tobacco mosaic virus (TMV) gels special problems arise from the nature of the diffraction pattern. The virus particles take up random orientation around their long axes and have only weak correlation with each other so that the diffraction pattern consists essentially of the cylindrical average of the square of the structure factor of a single particle. The diffraction pattern is limited to layer lines by the periodic nature of the virus but along each layer line the intensity is a continuous function which can be described as the sum of squares of a number of functions, the so-called **G** functions of Klug, Crick & Wyckoff (1958).

Each **G** function is related to a particular helical projection in the virus. Because of the high symmetry on any one layer line only three **G** functions are present out to 5 Å resolving power. At low resolution (< 12 Å) only one **G** function contributes to the observed intensity. If this condition pertains then

$$|F|^2=|G|^2=I,$$

where I is the measured intensity. Therefore, at low resolution the normal techniques of protein crystallography can be applied to the TMV diffraction pattern to produce an electron density map (Barrett *et al.*, 1971). At higher resolution the structure factor **F** can be recreated *if* the various **G** components can be separated and phases assigned to each. The present paper shows how this is in principle possible for any number of **G**'s and how in practice two **G**'s can be separated. The basis of the method is an extension of the method of isomorphous replacement. The results of this application to TMV have already been described (Holmes, Stubbs, Mandelkow & Gallwitz, 1975). That such a solution was possible in principle was pointed out by Holmes (1959). It is generally applicable to non-crystalline fibre diffraction problems, and to some crystalline fibres. Some model calculations are presented which indicate the reliability of the method. The application to TMV is discussed.

1. Theory of diffraction by a helix

Regular fibrous structures may be described as helices of repeating sub-units. Cochran, Crick & Vand (1952) showed that the structure factor of such a helix is

$$\mathbf{F}(R,\psi,l/c)=\sum_n \sum_j f_j J_n(2\pi R r_j) \times \exp i(-n\phi_j+2\pi l z_j/c) \exp in(\psi+\pi/2). \quad (1)$$

R, ψ and l/c are reciprocal-space cylindrical coordinates (l is an integer, the layer-line number, c is the repeat distance on the z axis). r_j, ϕ_j and z_j are the cylindrical coordinates of atom j in real space, which has scattering factor f_j.

J_n is the Bessel function of the first kind of order n. The summation over n is limited by the selection rule (determined by the helical symmetry)

$$l=tn+um \quad (2)$$

where m is an integer, and the helix has u subunits in t turns.

Klug, Crick & Wyckoff (1958) define the complex number

$$\mathbf{G}_{nl}(R)=\sum_j f_j J_n(2\pi R r_j) \exp i(-n\phi_j+2\pi l z_j/c) \quad (3)$$

in order to group together all terms containing Bessel functions of the same order.

Then

$$\mathbf{F}(R,\psi,l/c)=\sum_n \mathbf{G}_{nl}(R) \exp in(\psi+\pi/2). \quad (4)$$

Notation: since we are often interested in only one point (l,R) in the fibre diagram, the symbol $\mathbf{G}_n$ will be used for $\mathbf{G}_{nl}(R)$ where no confusion can arise.

Cylindrically averaging the intensity, *i.e.* over ψ, gives the intensity

$$I=\langle \mathbf{F}\mathbf{F}^* \rangle_\psi=\sum_n \mathbf{G}_n\mathbf{G}_n^* \quad (5)$$

as was shown by Waser (1955) and Franklin & Klug (1955). The phase problem is now to find the phase of each $\mathbf{G}_n$, and to separate the contribution of each $\mathbf{G}_n$. Because $\mathbf{G}_n$ contains only Bessel functions of order n, this problem is often referred to as that of 'separating Bessel functions'.

The problem is considerably simplified by the Bessel function property of having a negligible value until the argument approaches the order. Thus, for a particle of finite radius, only a small number of Bessel functions contribute to the diffraction pattern at a given resolution. For example, the diffraction pattern of TMV (maximum radius 90 Å) contains at most three Bessel function orders up to a resolution of about 7 Å. That is, equation (5) has at most three terms on the right-hand side, and the problem reduces to finding their relative magnitudes and phases. A procedure for doing this is given in Appendix *A*, using equations set up in the following section.

2. Isomorphous replacement as a general solution

In this procedure a heavy atom such as mercury is added to the native structure under investigation, without disturbing the structure in any other way. If $\mathbf{g}_n$ is the contribution to $\mathbf{G}_n$ by the heavy atom, that is, one term of the right-hand side of equation (3), we have

$$\mathbf{G} = \mathbf{G}_n + \mathbf{g}_n \qquad (6)$$

where the prime signifies a heavy-atom derivative. Let

$$\mathbf{G}_n = A_n + iB_n$$
$$\mathbf{g}_n = a_n + ib_n;$$

then from (5)

$$I = A_1^2 + B_1^2 + A_2^2 + B_2^2 + \ldots . \qquad (7a)$$

From (3) and (6)

$$I' = (A_1 + a_1)^2 + (B_1 + b_1)^2 + (A_2 + a_2)^2 + (B_2 + b_2)^2 \ldots . \qquad (7b)$$

Note that the subscripts 1,2 *etc.* are *labels* referring to different values of the Bessel function order n; they are not the actual values of n.

I and I' are observables. a_n and b_n may be calculated if the positions of the heavy atoms are known (see Holmes, Stubbs, Mandelkow & Gallwitz, 1975). We thus have a set of second-order equations in $2p$ unknowns, where p is the number of significant terms in equation (5). Equations similar to (7*a*) and (7*b*) may be used in crystallography, but with only the first two terms on the right-hand side. These are usually solved by a numerical search, minimizing the errors in equations (7*b*) or (more usually) the errors in the equations derived by taking the square root of (7*b*). However, the increased number of variables in this case makes a search procedure impractical, and can lead to errors. In Appendix *A*, an analytical procedure is given which finds not only the minimum errors, but all local minima. The procedure is in fact a minimization of the sum of the squares of the errors in equation (7*b*), subject to the constraint (7*a*). In other words, the errors in *derivative intensities* are minimized. This is not the same as the usual procedure of minimizing errors in *structure factors*, but is equally valid. The solutions found are the most probable (or locally most probable), and are preferred over 'best' solutions for reasons described below (§ 3).

3. Continuity as a constraint

In a non-crystalline fibre diffraction pattern, the layer-line intensities are not sampled, but vary continuously. In fact, the vectors $\mathbf{G}_n$ vary continuously, and this provides a valuable constraint on the phases of the points on a layer line taken as a whole. In its simplest form this is the principle of 'minimum wavelength' (Bragg & Perutz, 1952) by which the rate of fluctuation of $\mathbf{G}$ is restricted by the finite radius of TMV, as may be seen from equation (3).

This principle has been used to determine signs for the equator of TMV (Caspar, 1956) and barley stripe mosaic virus (Finch, 1965). The qualitative application of the constraint to single-Bessel-function regions of non-zero layer lines is typified in Fig. 1, which is an Argand diagram on which are marked all the most probable values of $\mathbf{G}$ for a region of the first layer line. A solid line connects points relating to adjacent points in R, and it is clear that $\mathbf{G}$ does indeed vary slowly and continuously.* The two points corresponding to $R=0{\cdot}056$ and $R=0{\cdot}065$ are not included in the line because they would cause discontinuities incompatible with the minimum-wavelength principle. However, in each of these cases, there is a second minimum in the error (see § 2), that is, the phase probability distribution is bimodal. This is the reason for determining all local error minima in Appendix *A*. These 'second most probable $\mathbf{G}$'s' are marked in the figure by crosses, and conform to the continuity requirement as well as any of the other points. In this way, $\mathbf{G}$ can be determined by a combination of isomorphous replacement and inspection having regard to continuity.

The procedure may readily be extended to a region containing p $\mathbf{G}$ terms, but there will be up to $2p$ error minima. Because of the difficulty in visualizing four or more dimensions (the real and imaginary parts of two or more $\mathbf{G}$ terms), the necessary decisions are best made on a quantitative basis, by a computer. To do this, certain assumptions must be made. For example, two possible $\mathbf{G}$ values may both conform to the minimum-wavelength principle, and one must be chosen as more probable. Furthermore, this probability should be weighed against probabilities determined by isomorphous replacement. In Appendix *B*, an al-

* Argand diagrams such as Fig. 1 have previously been used by A. Klug (private communication) to ensure continuity.

gorithm is given which places qualitative judgements on a quantitative basis, using the assumption that the second derivative of **G** with respect to R should be a minimum. However, it should be emphasized that the theory in this Appendix is by no means fully developed, and at this stage its primary virtue is that it permits a computer to determine very rapidly those values of **G** which best conform with our general ideas about continuity along a layer line.

4. Calculations for a model structure

We now have a procedure in which we determine a number of possible values of **G** for a point, each being locally the most probable, and then select one on the basis of continuity by considering a whole layer line. This procedure has been applied to a model data set in order to test the procedure, to find the number of heavy-atom derivatives required in practice given the expected level of error, and to determine the accuracy required in the positions of the heavy atoms. (Other heavy-atom parameters, such as occupancies and temperature factors, were not considered. Phases are not very sensitive to them at the resolution involved in this work, and these parameters cannot be satisfactorily refined, nor is it necessary to do so.)

The model used was part of the second layer line of a hypothetical structure having the same symmetry parameters as TMV. In Fig. 2 the real and imaginary parts of the first two **G** terms of the model are shown. In each experiment, heavy atoms were postulated at the positions given in Table 1, and **G** terms were calculated from equation (6) for the required number of derivatives. To simulate the type of errors we expect, errors in a Gaussian distribution were randomly added to the resulting structure factors [*i.e.* $\sqrt{I}$, see equation (5)]. We expect a certain degree of correlation between the errors, but the precise degree is not known. Somewhat arbitrarily, therefore, the errors were correlated by the formula

$$\varepsilon_n' = \tfrac{1}{4}(\varepsilon_{n-1} + 2\varepsilon_n + \varepsilon_{n+1})$$

where ε_n' is the error added to a point, and ε_{n-1}, ε_n and ε_{n+1} are the errors generated by the Gaussian error program for three neighbouring points. The final observations in real space were not noticeably sensitive to this step. The errors were normalized so that $E^2 = \langle \varepsilon^2 \rangle$ had a value consistent with observations from the TMV zero layer line.*

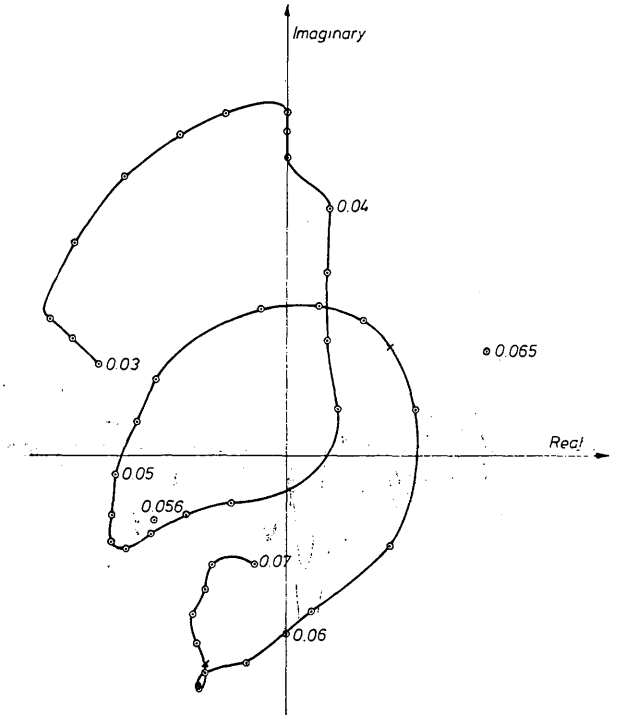

Fig. 1. Continuity of $\mathbf{G}_{-16,1}$ for TMV. Numbers are values of R. Each circle represents the most probable value of **G** on the Argand diagram. The crosses represent minor peaks in the probability distribution of points whose most probable **G**'s do not fit the general continuity in R.

Table 1. *Hypothetical heavy-atom sites*

The double derivative used sites 3 and 5. Units of f are arbitrary.

Site	f	r (Å)	ϕ(°)	z (Å)
1	35	40	18	7
2	36	50	0	0
3	40	58	13	19
4	32	82	15	13
5	28	72	8	10

Calculated intensities were then available and the phasing procedure described in §§ 2 and 3 was carried out. Both correct and incorrect heavy-atom positions were used (in the latter case, heavy atoms were all displaced in arbitrary directions). For example, Fig. 2 shows the original **G**'s, and those recalculated with five single-site derivatives and one double-site derivative,† phased from correct heavy-atom positions, and positions displaced 2 Å. If we turn our attention to the electron density maps, ϱ is given (Klug, Crick & Wyckoff, 1958) by

$$\varrho(r,\phi,z) = 1/c \sum_{l=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \gamma_{nl}(r) \exp\left[i(n\phi - 2\pi lz/c)\right],$$

where

$$\gamma_{n,l}(r) = \int_0^{\infty} \mathbf{G}_{n,l}(R) J_n(2\pi Rr) 2\pi R \mathrm{d}R.$$

* E can be estimated in the usual way from the zero layer line, but points of very low intensity near nodes in $G_{0,0}$ should not be included. This is because such points only occur on the centrosymmetric zero layer line, and they have errors rather larger than average.

† This combination was used because it is comparable with the situation in the TMV work, where five single and two double derivatives were used. A double derivative is usually one having sites in common with two single-site derivatives.

Clearly, our prime concern in model structure determinations is that the γ corresponding to the recalculated **G** should be recognizably that corresponding to the original. Therefore, γ was calculated for a number of situations. For the second Bessel function, the whole of **G** is not used to calculate γ, but only $G(R)$ where $R > R_0$. For $R < R_0$, no diffraction is possible for the higher-order G, because of the finite radius of the TMV particle and the properties of Bessel functions (see § 1). Therefore there is no need to separate Bessel functions. Furthermore, because at these radii the heavy-atom diffraction does not include higher orders, equations (7*b*) will be underdetermined and so large errors may occur if any attempt is made to separate the Bessel functions. (This is evident in Fig. 2, where large discrepancies between original and recalculated **G**'s are seen in $A_{-32,2}$ and $B_{-32,2}$ for $R < 0{\cdot}04$.)

Fig. 3 shows the imaginary part of the first γ term for the original **G**, for **G** recalculated using four single-site heavy-atom derivatives and for **G** recalculated using five single-site derivatives and one double-site derivative. Heavy-atom positions were assumed to be correctly known. It is clear that, given the expected level of errors, the method is capable of determining the γ terms with peaks positioned accurately to within about 2 Å. We see also that although increasing the number of derivatives does improve the accuracy, the improvement is small.

We next examine the effect of errors in the positions of the heavy atoms. Fig. 4 shows the same γ-term, this time determined with errors in the heavy-atom positions of 1, 2 and 4 Å. In each case it is compared with the correct γ. It appears that under these conditions errors of 1 Å can be tolerated, but errors of 2 Å or more will introduce serious errors into the structure. It is worth noting that in the refinement of heavy-atom positions referred to by Holmes, Stubbs, Mandelkow & Gallwitz (1975), movements of up to 2 Å from initially determined coordinates were sometimes observed. This suggests that even at this resolution, refinement of heavy-atom coordinates is desirable.

5. Application to TMV

Intensity data sets were available for native TMV, and seven derivatives having five independent sites. Details are in Table 2. Further information is given by Holmes, Stubbs, Mandelkow & Gallwitz (1975). The phasing procedure was carried out in two stages. Low-resolution data, to which only one Bessel function is expected to contribute, may be phased in a manner analogous to crystallographic isomorphous replacement, as mentioned in the *Introduction*.

At resolution between about 12 Å and 6·7 Å, the procedures described in the preceding paragraphs were used. Although the only certain guide to the satisfactory nature of the phasing is the interpretability of the resultant electron density map, there are two other checks available. One is that the **G** terms should be

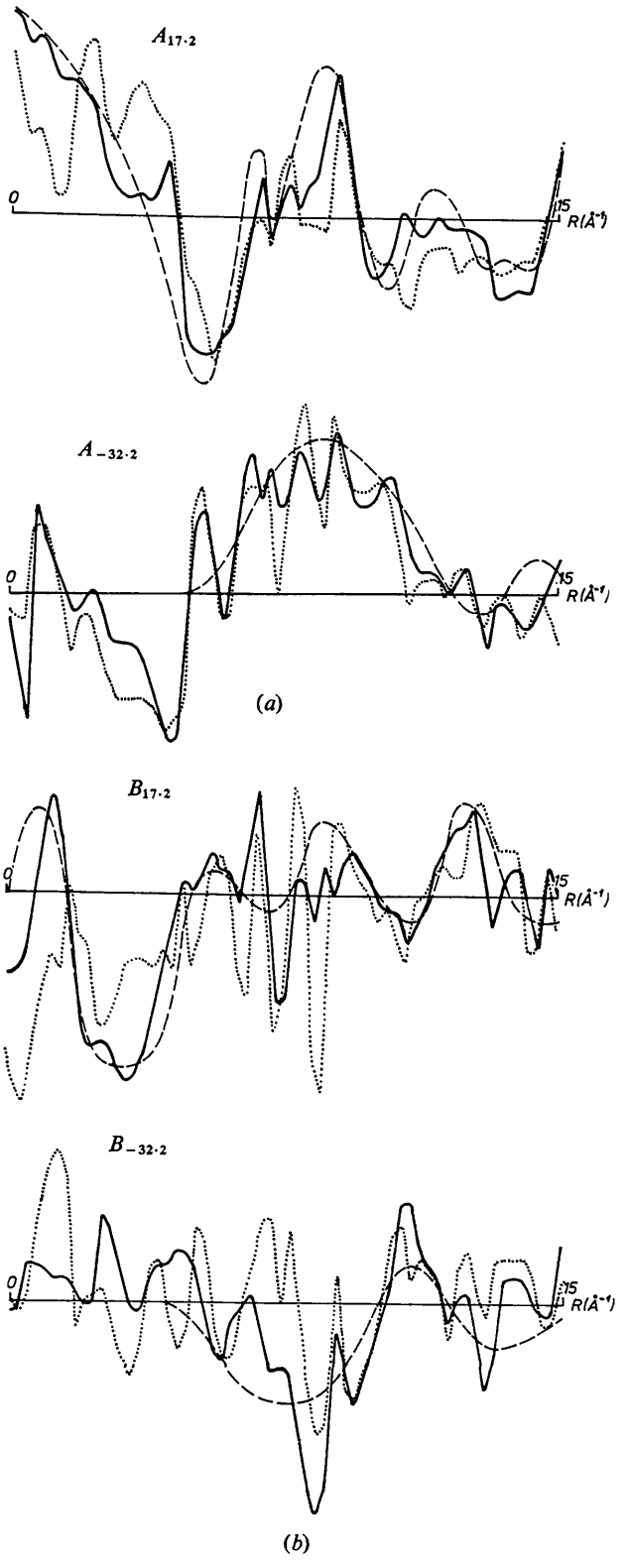

Fig. 2. **G** terms used as a model for testing the phasing procedures. Broken lines: original. Solid lines: recalculated using correctly positioned heavy atoms. Dotted lines: recalculated, heavy atoms displaced 2 Å. Real parts are shown in (*a*), imaginary in (*b*).

Table 2. *Heavy-atom derivatives*

Double derivatives contain sites 1 and 2, and 1 and 3. The SHIMS derivative also contains site 1.

Units of f are arbitrary (approximately electron/3).

Site	Ligand*	f	r (Å)	ϕ(°)	z (Å)
1	DMA	40	72	0	0
2	MMN	32	56·7	16·0	11·7
3	OsO_4	35	92·5	5·7	9·4
4	SHIMS-MMN	25	71·1	5·3	−1·3
5	Pb^{2+}	24	25·4	7·1	20·6

* SHIMS = sulphydryl imidoester attached to lysine 68.
DMA = dimercury acetic acid.
MMN = methyl mercury nitrate.

continuous. Although continuity has to some extent been imposed during the phasing, it is obviously not assured unless the phasing is sufficiently good. From Fig. 5(*a*), which shows the parts of the **G** terms for the first layer line of TMV, it can be seen that there are very few major departures from a reasonable degree of continuity. (A reasonable degree can be determined from the minimum-wavelength principle, which gives the minimum distance in R for one oscillation cycle as 0·01 Å^{-1} for a particle radius of 100 Å.) The other check is in the form of γ, and is closely related to the check on G. We know from the zero layer line that

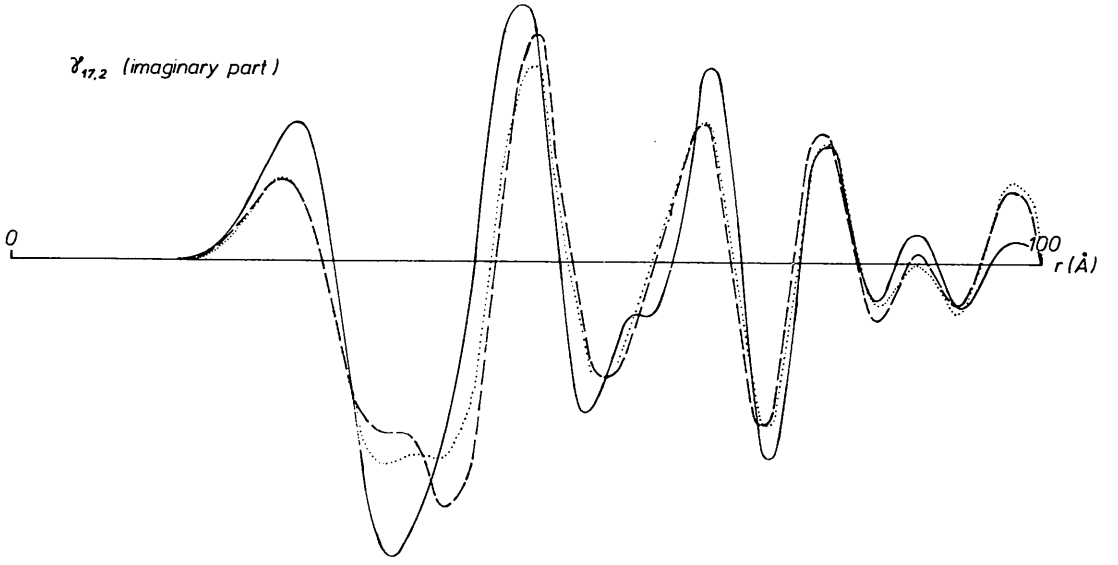

Fig. 3. Imaginary part of the first γ term for the model, showing the effect of extra heavy-atom derivatives. Solid lines: original. Dotted line: calculated using five single-site derivatives and one double. Broken line: calculated using four single-site derivatives (the algebraic minimum).

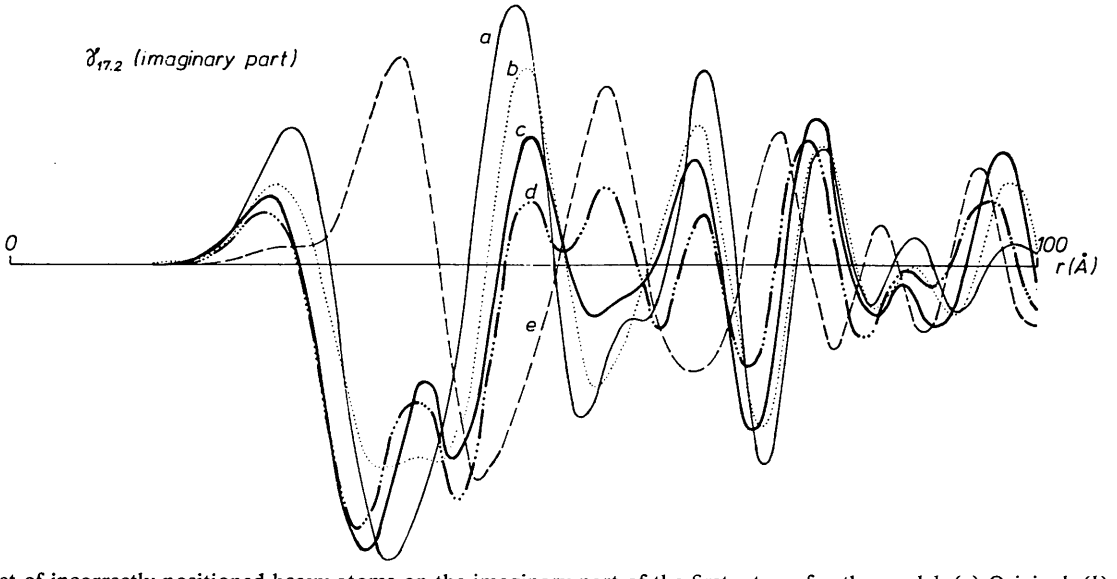

Fig. 4. Effect of incorrectly positioned heavy atoms on the imaginary part of the first γ term for the model. (*a*) Original. (*b*) Recalculated, heavy atoms correctly placed. (*c*) Recalculated, heavy atoms displaced 1 Å. (*d*) Recalculated, heavy atoms displaced 2 Å. (*e*) Recalculated, heavy atoms displaced 4 Å.

TMV is a hollow cylinder, of minimum radius about 20 Å and maximum radius about 90 Å (Franklin, 1956; Caspar, 1956). Therefore, we do not expect to see contributions to γ outside these radii. Fig. 5(*b*) shows the γ terms corresponding to the **G** terms in Fig. 5(*a*), and we see that most of the density is within the expected limits.

The Fourier synthesis which followed has been described by Holmes, Stubbs, Mandelkow & Gallwitz (1975). The high degree of interpretability of the electron density map is in itself confirmation of the power of these methods to determine sufficiently accurate phases. One further observation has been made, however, about the number of heavy-atom derivatives required to produce an accurate electron density map. An unpublished TMV map was produced using only six derivatives with four independent sites (the present set without the lead derivative). It was very similar to the five-site map, but there were one or two regions where interpretation was more difficult. In particular, the polypeptide chain in the five-site map is clear and unambiguous between residues 50 and 90. In the four-site map this density was broken, leading to an ambiguity in the chain tracing. Once again, therefore, it is clear that extra information from more derivatives than are algebraically required is small but useful.

6. Conclusion

Isomorphous replacement has been shown to be an adequate technique for separating terms of different Bessel-function order as well as phasing the terms in fibre diffraction patterns of TMV, and in model calculations. It appears that at the resolution of this work (6·7 Å) heavy-atom positions should be known within 1 Å. The minimum number of derivatives required is twice the number of terms to be separated and this could clearly present problems for structures having lower symmetry than TMV. However, experience with preparing derivatives is growing, and it is quite possible that a structure requiring three or four terms to be separated could be solved, given sufficiently good data.

I would like to thank Professor K. C. Holmes for his continued advice and encouragement during this work, and Mrs G. Eulefeld for her assistance with the figures. I would also like to thank Drs E. Mandelkow and J. Barrington Leigh for valuable discussions.

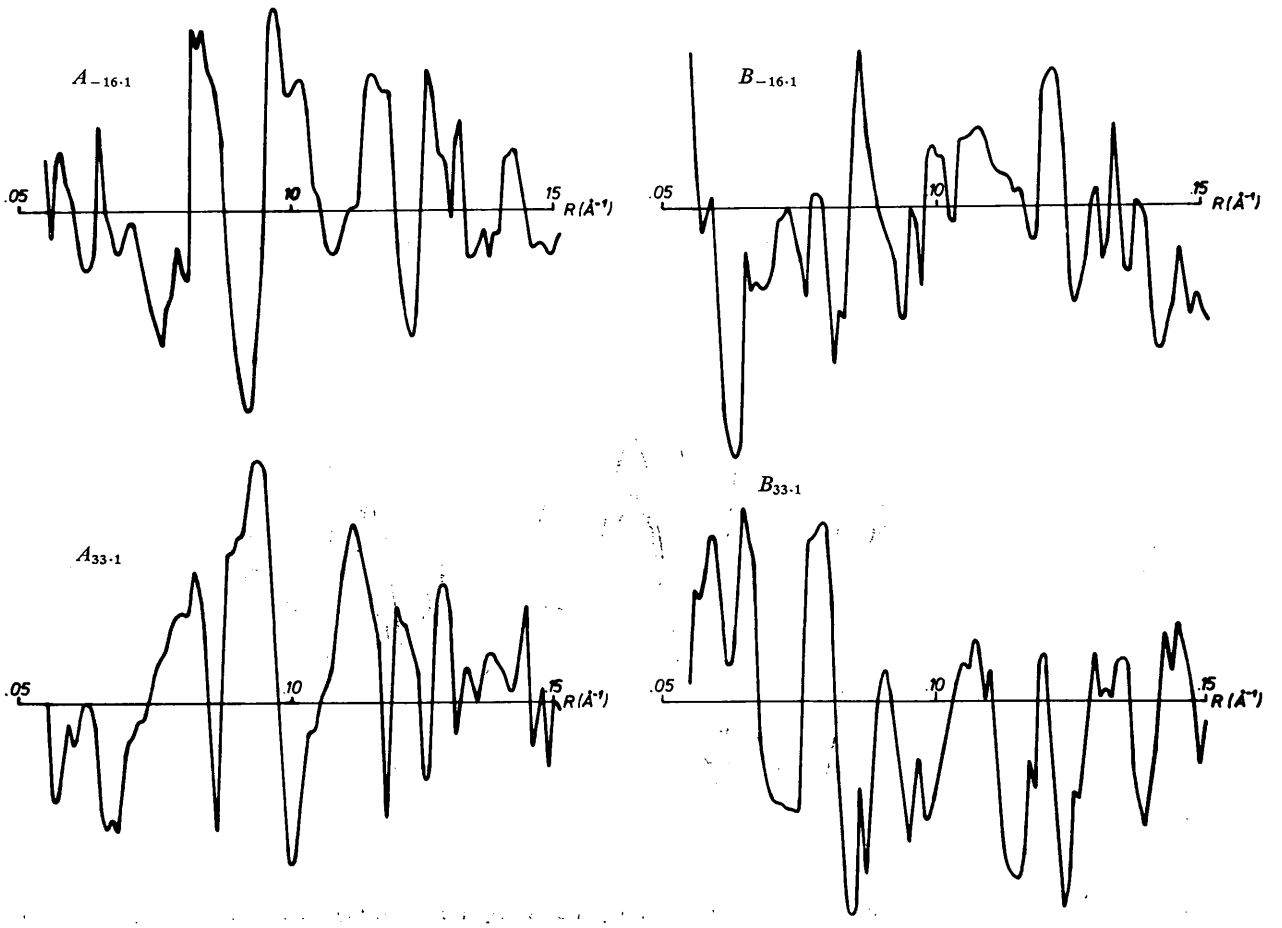

Fig. 5. (*a*) The parts of the **G** terms found for the first layer line of TMV.

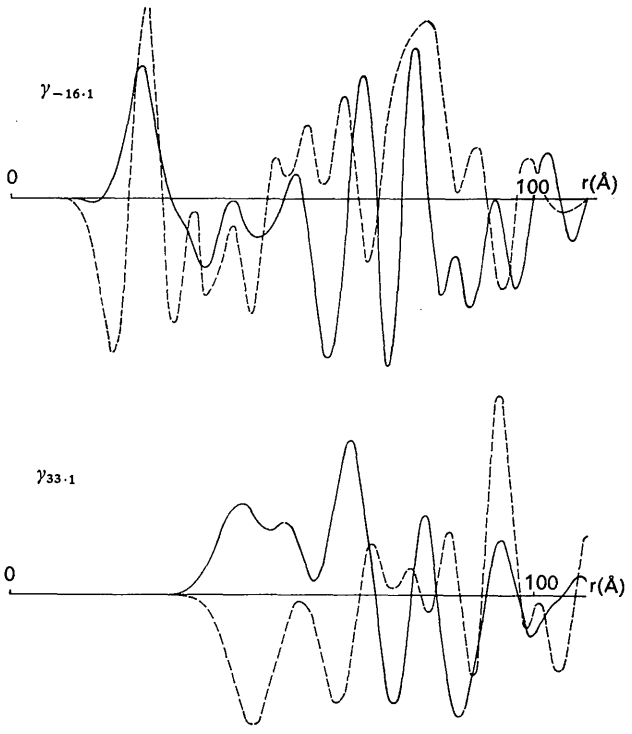

Fig. 5 (*cont.*). (*b*) The parts of the γ terms found for the first layer line of TMV. Solid lines represent real parts, broken lines imaginary parts. Structure is largely confined to radii between 20 and 100 Å.

APPENDIX A*

A least-squares procedure for minimizing the error in phases and Bessel-function term ratios

We have the equation

$$c_0^2=\sum_i x_i^2 \tag{A1}$$

and a set of equations

$$c_j^2=\sum_i (x_i+a_{ij})^2. \tag{A2}$$

These correspond to equations (7*a*) and (7*b*). Each x corresponds to either the real or the imaginary part of some $\mathbf{G}_n$. a_{ij} is the component of the diffraction from heavy atom j corresponding to x_i.

This overdetermined set of second-order equations could be solved by a numerical search procedure, seeking to minimize the error in equation (A2) (and possibly A1), but the procedure described here was faster by a factor of 50 on a Sigma 2 computer, as well as being slightly more accurate. We make use of the fact that errors in equation (A1) are usually much smaller than those in (A2). This is particularly so for very carefully collected data, in which errors are caused mainly by lack of isomorphism. We then consider only the errors ε_j in the derivative intensities c_j^2 and from (A2) obtain

$$c_j^2=\sum_i (x_i+a_{ij})^2+2\varepsilon_j \tag{A3}$$

and, substituting (A1) in (A3),

$$\sum a_{ij}x_i=\tfrac{1}{2}(c_j^2-c_0^2-\sum_i a_{ij}^2)+\varepsilon_j$$

or, in matrix notation,

$$A\mathbf{x}=\mathbf{b}+\boldsymbol{\varepsilon} \tag{A4}$$

where

$$b_j=\tfrac{1}{2}(c_j^2-c_0^2-\sum_i a_{ij}^2).$$

(A4) is a set of linear equations, but any solution is subject to the constraint (A1) used in deriving the set.

We shall assume that all the equations (A4) have equal weight. In practice this is achieved by dividing each equation by the expectation value of ε_j, which is derived below.

We are interested in minimizing the sum of error squares, $\boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon}$ (the superscript T denotes the transpose of a matrix or vector), subject to the constraint (A1), and we want to find *all* local minima.

Consider

$$\boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon}=\mathbf{b}^T\mathbf{b}-2\mathbf{b}^TA\mathbf{x}+\mathbf{x}^TA^TA\mathbf{x}. \tag{A5}$$

This is the equation of a family of ellipsoids, each determined by a value of $\boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon}$. They are illustrated in Fig. 6, together with the spheroid represented by equation (A1), represented as a two-dimensional case. It is clear that any solution $\mathbf{x}$ to our problem is a point where an ellipsoid and the spheroid are tangential, and thus have parallel normals. Now the normal to any ellipsoid (A5) is

$$M\mathbf{x}-A^T\mathbf{b}$$

where $M=A^TA$.

Therefore

$$M\mathbf{x}-A^T\mathbf{b}=\mu\mathbf{x}. \tag{A6}$$

μ is a scalar, not yet known, associated with $\mathbf{x}$, and we have to find those values of μ corresponding to $\mathbf{x}^T\mathbf{x}=c_o^2$. Clearly

$$(M-\mu I)\mathbf{x}=A^T\mathbf{b} \tag{A7}$$

(I is the identity matrix). Now, let E contain the eigenvectors of M such that

$$E^TME=\Lambda, \quad E^TE=I$$

(Λ is the diagonal matrix of eigenvalues, whose ith element is λ_i).

Let

$$\mathbf{v}=E^T\mathbf{x} \quad (\mathbf{x}=E\mathbf{v}). \tag{A8}$$

Pre-multiplying (A7) by E^T,

$$E^T(M-\mu I)E\mathbf{v}=E^TA^T\mathbf{b}=\boldsymbol{\omega} \text{ (this defines } \boldsymbol{\omega}).$$

Therefore

$$(\Lambda-\mu I)\mathbf{v}=\boldsymbol{\omega}.$$

* By R. Diamond and G. J. Stubbs.

The ith component of $\mathbf{v}$ is thus seen to be

$$v_i = \frac{\omega_i}{\lambda_i - \mu}$$

Imposing the condition (A1) gives us

$$c_0^2 = \mathbf{x}^T\mathbf{x} = \mathbf{v}^T\mathbf{v} = \sum \left(\frac{\omega_i}{\lambda_i - \mu}\right)^2 . \qquad \text{(A9)}$$

From this we may determine values of μ. For an n-dimensional problem there will clearly be between 2 and $2n$ solutions of (A9). These may be found in many ways. The method used in this work was a Newton–Raphson iteration applied to the octic in μ derived from (A9). Alternatively, Newton–Raphson iteration may be applied directly to (A9), starting from any point μ close enough to λ_i to ensure that

$$\sum \left(\frac{\omega_i}{\lambda_i - \mu}\right)^2 - c_0^2 > 0.$$

In this case, it will be immediately obvious whether a real pair of solutions exists between λ_i and λ_{i+1}, as the sequence of iterates is monotonic and bounded by λ and λ_{i+1} if and only if such a pair exists.

Once the solutions μ have been found, a set of values of $\mathbf{x}$ may be derived from (A8). These will include maxima, minima and saddle points.

The nature of the stationary point $\mathbf{x}$ is determined by considering a general point $\mathbf{x}+\delta\mathbf{x}$.

Let

$$\delta\varepsilon^2 = \varepsilon_g^T\varepsilon_g - \varepsilon_s^T\varepsilon_s$$

(subscripts g and s refer to general and stationary points).

From (A5) and (A6),

$$\delta\varepsilon^2 = \delta\mathbf{x}^T M\delta\mathbf{x} + 2\mu\delta\mathbf{x}^T\mathbf{x}.$$

If $(\mathbf{x}^T+\delta\mathbf{x}^T)(\mathbf{x}+\delta\mathbf{x}) = r^2$

$$\delta\varepsilon^2 = \delta\mathbf{x}^T(M-\mu I)\delta\mathbf{x} + \mu(r^2 - c_0^2). \qquad \text{(A10)}$$

Now let $R\delta\mathbf{u} = \delta\mathbf{x}$, where R is any orthogonal matrix such that $R^T\mathbf{x}$ is a vector with all components zero except the last. [R may be constructed from simple rotation matrices (Rollett, 1965).]

Let

$$S_n = R^T(M-\mu I)R.$$

Then

$$\delta\varepsilon^2 = \delta\mathbf{u}^T S_n\delta\mathbf{u} + \mu(r^2 - c_0^2). \qquad \text{(A11)}$$

We are concerned with the sign of $\delta\varepsilon^2$ for perturbations in the hyper-surface of the spheroid, so the second term in (A11) may be dropped.

Let $\delta\mathbf{u}_{n-1}$ be the $n-1$-dimensional vector containing all but the last element of $\delta\mathbf{u}$, and S_{n-1} the corresponding submatrix of S_n. Then in a domain around $\mathbf{x}$, $\delta\varepsilon^2$ has the same sign as $\delta\mathbf{u}_{n-1}^T S_{n-1}\delta\mathbf{u}_{n-1}$. This sign depends upon the eigenvalues of S_{n-1}, so $\mathbf{x}$ may be characterized as a minimum if they are all positive, a maximum if they are all negative, and otherwise as a saddle point.

Note on weights: The expectation value E of ε [see equation (A3)] could be estimated by studying a centric layer line (containing only a J_0 term), but E is a function of F_H, and has been found difficult to estimate reliably from the limited number of data available. In contrast, the expectation value E_F of the error ε_F in F has been found here (in the single-Bessel-function region of the fibre diagram) and in a crystallographic case (Stubbs, 1972) to be independent of F. Therefore, we have estimated E_F from the zero layer line, and calculated E as follows:

$$\varepsilon = \varepsilon_F(\varepsilon_F + 2F) \qquad \text{(A12)}$$

$$E^2 = \langle\varepsilon^2\rangle \text{ and } E_F^2 = \langle\varepsilon_F^2\rangle.$$

Squaring (A12) and taking mean values,

$$\langle\varepsilon^2\rangle = \langle\varepsilon_F^4\rangle + 4F\langle\varepsilon_F^3\rangle + 4F^2\langle\varepsilon_F^2\rangle.$$

We use the fact that ε_F has approximately a Gaussian distribution to obtain the relationship

$$E^2 = 3E_F^4 + 4F^2E_F^2. \qquad \text{(A13)}$$

Equation (A13) has been confirmed experimentally by Stubbs (1972) in a crystallographic case where there was sufficient centric data to estimate E and E_F independently.

We note that the procedure described in this Appendix could be applied to the crystallographic phase problem, in which equation (A1) has only two terms. Most probable phases have often been used in refinement procedures [for example, by Kraut, Sieker, High & Freer (1962)] and in such cases, the direct determination of the phase by this method will be very much faster than a numerical procedure. Note that the most probable phase corresponds to the smallest μ [as may be deduced from equation (A10), remembering that in this case $\delta\varepsilon^2$ must always be positive] and so characterization of the stationary points is not required.

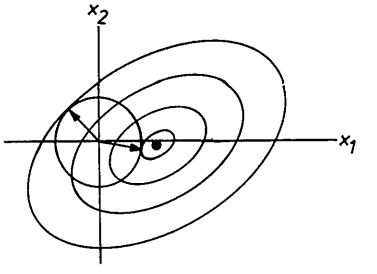

Fig. 6. The family of ellipsoids, centred on the heavy dot, representing isometric values of the residual $\varepsilon^T\varepsilon$ in equation (A5). The circle represents equation (A1), the constraint on the solution vector. Two possible values of $\mathbf{x}$ are shown, each being a local minimum in the residual subject to the constraint.

APPENDIX B

The use of continuity of diffraction in phase determination

Continuity along a layer line stems from the finite size of a TMV particle which necessitates that the components of **G** do not fluctuate too rapidly. We can see this by considering a component

$$A_{n,l}(R)=\sum_j f_j J_n(2\pi R r_j)\cos\theta_j \qquad \text{(B1)}$$

($\cos\theta_j$ is a phase factor independent of R).

Since any one $J_n(2\pi Rr_j)$ does not fluctuate rapidly, and r_j has a strictly limited upper bound, $A_{n,l}(R)$ is correlated in R to an extent related to the value of r_j. To express this correlation and use it in phase determination would be the proper use of continuity, but this would involve solving the phase problem for all the points on a layer line simultaneously, which presents serious difficulties in computer storage. Instead a method is presented below of finding the smoothest function by an iterative procedure. Smoothness or correlation is maximized by minimizing the high-order derivatives of a function (in this case, A with respect to R). For example, writing out the kth term of the Taylor expansion for A we get:

$$(R-R_0)^k/k!\sum_j[(2\pi r_j)^k J^{(k)}(2\pi Rr_j)f_j\cos\theta_j]$$

[(k) signifies differentiation by the argument k times].

Now, $J^{(k)}(z)$ is of the same order of magnitude as $J(z)$, so the ratio of term k to term $(k+1)$ is approximately

$$(k+1)/2\pi r_0(R-R_0)$$

where r_0 is the mean radius of the diffracting object. For $R-R_0=0{\cdot}0015$, and $r_0=65$, the first three terms have relative magnitudes

$$1:0{\cdot}6:0{\cdot}2.$$

Thus, given a choice of two sets $\{A\}$, we choose the set in which the second derivatives are minimized. (In other words, the curvature in Fig. 1 is minimized.) This procedure has the advantage, as will be shown below, that it can be applied to a layer line in a single pass, considering only three points at a time.

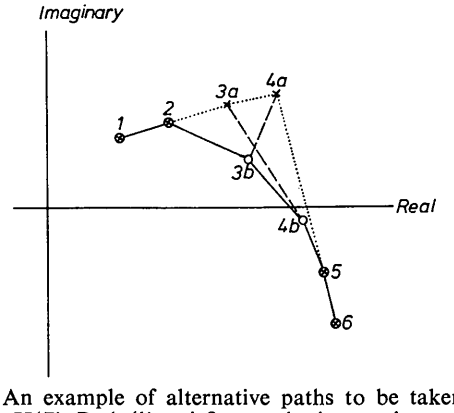

Fig. 7. An example of alternative paths to be taken by the vector $\mathbf{H}(R)$. Path (b) satisfies continuity requirements better than the other three possibilities.

If we have a sequence of closely spaced equidistant points $R_1, R_2, \ldots$, we may estimate the probability of the value of A_{i+1}:

$$P(A_{i+1})=k\exp-[(A_{i+1}-2A_i+A_{i-1})^2/2E^2]. \qquad \text{(B2)}$$

An implicit assumption is that the second derivatives of A have a Gaussian distribution about zero, standard deviation E. While this is not strictly valid (the third and first terms of the Taylor series are slightly correlated), it has been found to serve well enough for practical purposes.

E^2 can be estimated from the zero layer line, or any region such as the single-Bessel-function region where phases are already known with reasonable certainty. (A continuity diagram such as Fig. 1 would be a suitable source.) In practice, since only a small number of possible values for A_{i+1} are considered, the finally selected sequence is very insensitive to E.

Now let $\mathbf{H}(R)$ be a multidimensional vector having the components of all the **G** factors contributing to diffraction at a point in reciprocal space. By repeated use of (B2) we may choose between different sets $\{\mathbf{H}(R_i)\}$. If we have a number of possible values for each $\mathbf{H}(R_i)$, it is not necessary to consider every possible set. We consider the values of i in turn. It is necessary to consider every possible value of $\mathbf{H}(R_i)$ and $\mathbf{H}(R_{i-1})$, but values with index less than i do not affect the calculation of subsequent probabilities. The algorithm for calculating the most probable $\{\mathbf{H}(R_i)\}$ is as follows: [Notation: h_{ij} is the jth possible value for $\mathbf{H}(R_i)$]. In store we have a number of sequences $\{h_{1j}\ldots h_{ij}\}$, each terminating in a different pair $\{h_{i-1j},h_{ij}\}$. We consider every pair $\{h_{ij},h_{i+1j}\}$ and calculate from (B2) the probability that each of the sequences in store terminating in h_{ij} continues with h_{i+1j}. The sequence with the highest probability forms the beginning of the new sequence $\{h_{1j}\ldots h_{ij},h_{i+1j}\}$. When all values of i are exhausted, the sequence with the highest joint probability [*i.e.* $\prod_{i=3}^{n} P(h_{ij})$] is accepted.

A two-dimensional example is given in Fig. 7. Six points, closely spaced in R, are shown. (For interpretation of the diagram see Fig. 1.) There are alternatives available for points 3 and 4, so four pathways are possible. The following sequences are stored.

Step 1: $\{1,2\}$
Step 2: $\{1,2,3a\}$ and $\{1,2,3b\}$
Step 3: $\{1,2,3a,4a\}$ $\{1,2,3b,4a\}$
$\{1,2,3a,4b\}$ $\{1,2,3b,4b\}$

(at this stage $\{1,2,3a,4a\}$ has a much higher probability than the others, but all are retained)

Step 4: $\{1,2,3a,4a,5\}$ and $\{1,2,3b,4b,5\}$.

Two sequences have now been dropped, since no

possible point 6 can make them more probable than the sequences retained. The *b* sequence is now more probable. In step 5, when point 6 is known, only the *b* sequence is retained.

If the elements h_{ij} have different probabilities apart from continuity considerations (which will normally be the case, if they have initially been found by some such method as isomorphous replacement), these initial probabilities can be included at both the sequence-extension stage and the final selection stage.

References

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

BRAGG, W. L. & PERUTZ, M. F. (1952). *Proc. Roy. Soc.* A**213**, 425–435.

CASPAR, D. L. D. (1956). *Nature, Lond.* **177**, 928.

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

FINCH, J. T. (1965). *J. Mol. Biol.* **12**, 612–619.

FRANKLIN, R. E. (1956). *Nature, Lond.* **177**, 928–930.

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

HOLMES, K. C. (1959). *X-ray Diffraction Studies on Tobacco Mosaic Virus and Related Substances.* Thesis, Univ. of London.

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

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

KRAUT, J., SIEKER, L. C., HIGH, D. F. & FREER, S. T. (1962). *Proc. Natl. Acad. Sci. U.S.* **48**, 1417–1424.

ROLLETT, J. S. (1965). *Computing Methods in Crystallography.* Oxford: Pergamon Press.

STUBBS, G. J. (1972). *Structural Studies of Crystalline Proteins.* Thesis, Univ. of Oxford.

WASER, J. (1955). *Acta Cryst.* **8**, 142–150.

Acta Cryst. (1975). A**31**, 718

Self-Crystallizing Molecular Models. IV. Revision and Conclusion

BY TARO KIHARA

Department of Physics, Faculty of Science, University of Tokyo, Tokyo, Japan

(*Received* 31 *March* 1975; *accepted* 23 *April* 1975)

Molecular models with magnetic multipoles, which were invented by the author for the purpose of simulating crystal structures, are revised. The models of the new type are made of Ba ferrite magnets and plastic pieces only, no Mn–Zn ferrite being used. The crystal structures of SiF_4 and α-CF_4, as well as those of UCl_6 and WCl_6 are represented by use of different species of molecular models.

Introduction

The purpose of this series of papers (Kihara, 1963, 1966, 1970) is to explain the structures of molecular crystals in terms of the shapes of the molecules as well as the intermolecular force. We treat rigid nonpolar molecules with no power to form hydrogen bonds.

If the molecules do not possess any appreciable electric multipoles, the crystal structures are governed by the principle of closest packing of the molecules. The orthorhombic crystals of Cl_2, Br_2 and I_2, and the cubic crystals of SiI_4, GeI_4, SnI_4, *etc.* are examples of such structures.

If, on the other hand, the molecules have sufficiently strong electric multipoles, the electrostatic interaction often governs the crystal structure. In such cases, the crystal structure does not necessarily correspond to the closest packing of the molecules.

The electrostatic multipolar interaction between molecules can be represented by use of molecular models with magnetic multipoles. A structure into which these models are assembled will simulate the actual crystal structure.

In Part II (1966), eight types of molecular models made of barium ferrite magnets were given. In Part III (1970), models consisting of barium ferrite magnets

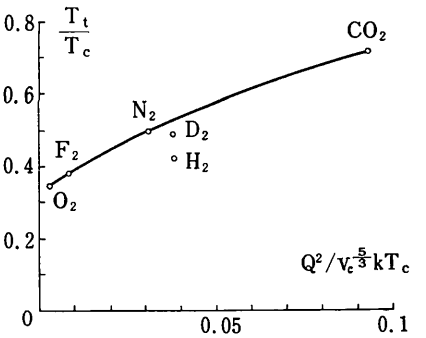

Fig. 1. Ratio of the triple-point to the critical-point temperatures as a function of the dimensionless quadrupolar interaction.