When the variables are uncorrelated, (51) gives weights inversely proportional to the corresponding variances.

For the case of just two correlated variables

$$1/w_i = \sigma^2(l_i) - \text{cov}\,(l_1, l_2)\,. \qquad (52)$$

References

ARLEY, N. & BUCH, K. R. (1950). *Introduction to the Theory of Probability and Statistics.* New York: John Wiley.

BUSING, W. R. & LEVY, H. A. (1958). *Acta Cryst.* **11**, 450.

BUSING, W. R., MARTIN, K. O. & LEVY, H. A. (1964). *ORFFE, A Fortran Crystallographic Function and Error Program.* Oak Ridge National Laboratory, Oak Ridge, Tennessee.

CRUICKSHANK, D. W. J. & ROBERTSON, A. P. (1953). *Acta Cryst.* **6**, 698.

DARLOW,S.F. (1960). *Acta Cryst.* **13**, 683.

HAMILTON, W. C. (1964). *Statistics in Physical Science.* New York: Ronald Press.

International Tables for X-ray Crystallography (1959). Vol. II. Birmingham: Kynoch Press.

TEMPLETON, D. H. (1959). *Acta Cryst.* **12**, 771.

TEMPLETON, D. H. (1960). *Z. Kristallogr.* **113**, 234.

WASER, J. (1955). *Acta Cryst.* **8**, 731.

A Description of Various Rotation Function Programs

BY PATRICK TOLLIN AND MICHAEL G. ROSSMANN

Department of Biological Sciences, Purdue University, Lafayette, Indiana 47907, U.S.A.

Various closely related programs for the calculation of the rotation function are described. The latter explores systematically the amount of overlap between two differently oriented Patterson syntheses, and can be used to relate similar molecules or structures in the same or different crystals. The calculations require only the intensities rather than the Patterson sections. It is shown that (i) neglecting all but 10 % of the largest intensities for one of the structures and (ii) construction of a table of the transform G, of the spherical volume within which the Patterson functions are being compared, sampled in a $5 \times 5 \times 5$ grid within the reciprocal unit cell, gives considerable improvement in computing time without excess loss of accuracy. The effect of premature truncation or coarseness of the G table is discussed, together with other considerations which are important in the successful application of this technique.

1. Introduction

We shall describe the flow diagram of various closely related programs for the calculation of the rotation function (Rossmann & Blow, 1962) (RB). Even with existing fast computers the time involved in exploring the three rotation angles at reasonable intervals is formidable and a number of techniques are here presented which significantly improve the speed of the computations. We also discuss strategic considerations required in the application of these techniques to various types of problem.

The rotation function is defined (RB) as

$$R = \int_U P_2(\mathbf{x}_2)\,.\,P_1(\mathbf{x}_1)\,d\mathbf{x}_1\,. \qquad (1)$$

It measures the degree of coincidence when the Patterson function P_1 is rotated on the Patterson function P_2. Any point $\mathbf{x}_1$ in P_1 is related to any other point $\mathbf{x}_2$ in P_2 through the rotation matrix [C] by the relationship

$$\mathbf{x}_2 = [\mathbf{C}]\mathbf{x}_1\,.$$

The above integral (1) can be shown (RB) to reduce to the double summation

$$R = \sum_{\mathbf{p}} |F_{\mathbf{p}}|^2 \left\{ \sum_{\mathbf{h}} |F_{\mathbf{h}}|^2\, G_{\mathbf{h},\mathbf{h}'} \right\} \qquad (2)$$

where $|F_{\mathbf{p}}|$ and $|F_{\mathbf{h}}|$ are the structure amplitudes corresponding to the Patterson functions P_2 and P_1 respectively. $G_{\mathbf{h},\mathbf{h}'}$ is an interference function whose magnitude depends on the reciprocal lattice vectors $\mathbf{h}$ and $\mathbf{h}'$ as well as the volume U within which the integral (1) is evaluated. The non-integral reciprocal lattice vector $\mathbf{h}'$ is given by

$$\mathbf{h}' = \mathbf{p}[\mathbf{C}]\,.$$

The rotation function is particularly useful for the following problems:

(a) To determine the relative orientation of identical or similar rigid chemical groups in two different crystals. P_1 and P_2 must then represent the Patterson functions of the two crystals.

(b) To determine the orientation of a known rigid group in a molecular crystal. Here P_1 is the Patterson function of the unknown crystal, while P_2 is the Patterson function calculated from a model of rigid group in a known orientation.

(c) To determine the relative orientation of identical or similar groups of molecules within the same crystallographic asymmetric unit. Now P_1 and P_2 both represent the same Patterson function of the unknown crystal.

In all cases the integration is performed over the volume U equal to the volume around the origin of the Patter-

son functions within which P_1 and P_2 are expected to show similarity. We have invariably chosen U to be a sphere of radius r_0; $G_{\mathbf{h},\mathbf{h}'}$ is then determined entirely by the magnitude of the vector $\mathbf{H} = \mathbf{h} + \mathbf{h}'$, and is then given by the expression

$$G_{\mathbf{h},\mathbf{h}'} = \frac{3(\sin 2\pi H \cdot r_0 - 2\pi H r_0 \cos 2\pi H r_0)}{(2\pi H r_0)^3}.$$

It has previously been observed (RB) that $G_{\mathbf{h},\mathbf{h}'}$ has a large value only if H is small. Thus the inner summation $\{\sum_{\mathbf{h}} |F_{\mathbf{h}}|^2 G_{\mathbf{h},\mathbf{h},}\}$ in (2) need only be performed over those points $\mathbf{h}$ which are close to the non-integral point $-\mathbf{h}'$. In order to maintain the exact rotation symmetry (Tollin, Main & Rossmann, 1966) it is important to select a set of symmetrically disposed points around $\mathbf{h}$. The number, n, of points which need be included in this summation will depend on the magnitude of the radius, r_0, and the reciprocal unit cell dimensions of the crystal P_2. A discussion of the best value of n is given in §5, but for most purposes 27 points seems satisfactory.

2. The large term program*

Examination of equation (2) indicates that R will have a large value when large values of $|F_{\mathbf{h}}|^2$ and $|F_{\mathbf{p}}|^2$ are associated by large values of $G_{\mathbf{h},\mathbf{h}'}$. Since R is approximately proportional to $|F|^4$ it seems unlikely that the sum of the products of many small intensities will outweigh the contribution of a few large products. Also, since the time to compute R is directly proportional to the number of $|F_{\mathbf{p}}|^2$ values, and is only slightly af-

---

* The idea that using only the largest $|F|^2$ values would give useful results was arrived at independently by Dr D. M. Blow.
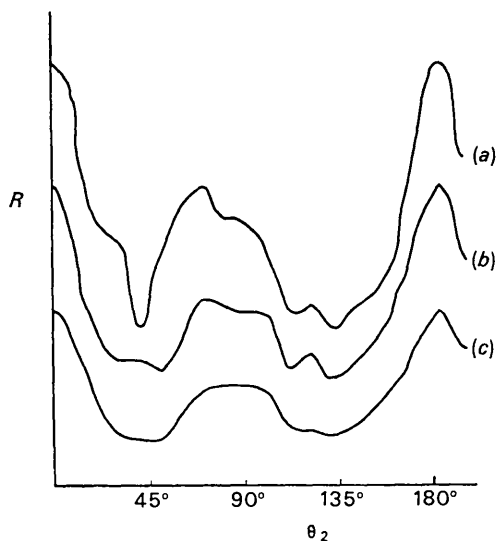

Fig. 1. A line through the monoclinic insulin rotation function in order to compare results using (a) full data, (b) only the 50 largest terms for the $|F_{\mathbf{p}}|^2$, and (c) the 50 largest terms for the $|F_{\mathbf{h}}|^2$ and $|F_{\mathbf{p}}|^2$ set of intensities.

fected by the number of $|F_{\mathbf{h}}|^2$, a modification of the general program was written which uses a small number of the largest $|F_{\mathbf{p}}|^2$ values. The above argument has been tested by computing the rotation function of monoclinic insulin on itself along the line $\theta_1 = 90°$, $0 \le \theta_2 \le 180°$, $\theta_3 = 90°$, where θ_1, θ_2, θ_3 are the Eulerian angles as defined by RB. This particular line is known to have significant peaks along it (Hodgkin, Dodson, Coller & Rossmann, unpublished). The results obtained are shown in Fig. 1. The top curve shows the form of R when all the data were used for both $|F_{\mathbf{p}}|^2$ and $|F_{\mathbf{h}}|^2$. The middle curve shows the form of the rotation function when only the 50 largest $|F_{\mathbf{p}}|^2$ values and all $|F_{\mathbf{h}}|^2$ data had been employed. The bottom curve shows the result of using only the 50 largest intensities of both the h (or first) and p (or second) crystals. Although using only the largest $|F_{\mathbf{h}}|^2$ intensities does not make a great saving in time of computation for our program on a $32K$ memory store machine, the fact that useful results were obtained might be helpful in constructing a fast rotation function program for a smaller computer.

A further indication that this approximation is justified was obtained in the comparison of seal and sperm whale myoglobin (Tollin & Scouloudi, 1966). In the latter example the time per point, using all the reflection data, was 24 seconds, while when using only the 47 largest $|F_{\mathbf{p}}|^2$ values the time per point reduced to 1·8 seconds. These times refer to a Fortran II program compiled and executed on an IBM 7094 computer. The latter has a cycle time of 2 μsec, requires on the average about 10 cycles per floating point operation, and has 32768 words of memory.

3. The fine mesh program

Examination of the block diagram of the general program (Fig. 2) shows that most of the time is spent in calculating the magnitude of $\mathbf{H}$ in the innermost loop. Much of this time can be saved by noting that the point $-\mathbf{h}'$ can be written as

$$h_1 + \Delta h_1', h_2 + \Delta h_2', h_3 + \Delta h_3',$$

where the nearest integral reciprocal lattice point to $-\mathbf{h}'$ has coordinates (h_1, h_2, h_3). A three-dimensional table can be constructed of values of $G_{\mathbf{h},\mathbf{h}'}$ with respect to a given integral point $\mathbf{h}$ and all values of $\Delta h_1'$, $\Delta h_2'$, $\Delta h_3'$, laid out as a fine grid within the reciprocal unit cell. In order to keep the size of this table within bounds each reciprocal length was divided into only five equal parts, and the summation over n was fixed at 27 points. Hence this table contained 27×125 entries. The computation of the inner loop is therefore reduced to determining $\Delta h_1'$, $\Delta h_2'$, $\Delta h_3'$ and then looking up the corresponding 27 values of $G_{\mathbf{h},\mathbf{h}'}$.

Fig. 3 shows the result of using this program for monoclinic insulin. The time per point, using this technique in the seal versus whale comparison was now only 1·0 sec when only the 47 largest $|F_{\mathbf{p}}|^2$ values were used.

4. The block diagram of the program

The block diagram of the general and large term programs is given in Fig.2(a). The modifications to produce the 'fine mesh' program are shown in Fig.2(b). A number of explanatory notes are given below. Each step in the program is represented by a block in the diagram.

The general and large-term programs

Step 1. The cell data for both crystals, the Laue symmetry of reciprocal space, the limits of the rotation angles, the value of r_0 (the radius of the sphere of Patterson matching), and the maximum and minimum $\sin\theta/\lambda$ for the reflections to be used are read into the computer.

Step 2. A table of $G_{h,h'}$ is constructed as a function of $(Hr_0)^2$. Values of G are listed for 400 equally spaced values of $(Hr_0)^2$ between $0 \le Hr_0 \le 2\cdot0$.

Step 3. The $|F|^2$ values for both crystals are read into the computer and a hemisphere in reciprocal space is generated using their Laue symmetries. Reflections outside the given limits of $(\sin\theta/\lambda)$ are rejected. The position of any $|F|^2$ value in the store is a function of its Miller indices alone in order to give immediate random access to each structure amplitude. $|F|^2$ values with the same Miller indices for both the first and second crystals are packed into the same store location to save space. However in the 'large-terms' program a separate table lists the Miller indices and $|F_p|^2$ values of the largest terms. This gives the twofold advantage of avoiding not only generating Miller indices in step 7, of which approximately one-third will necessarily be outside the permitted limits of $(\sin\theta/\lambda)$, but also the unpacking procedure while looking up each $|F_h|^2$ value.

Step 4. As indicated by RB, if x_1 and x_2 represent fractional coordinates with respect to the unit cell axes of each crystal, we may express the rotation matrix [C] as the triple product [α][ϱ][β], where [β] converts x_1 to an orthogonal system and [α] converts from an orthogonal system back to fractional coordinates. Matrix [β] is then a function of the rotation angles alone. Matrices [α] and [β] are set up here.

Step 5. The values of the rotation angles are generated within the range defined by the data read in step 1. Our present programs use only Eulerian angles as these lead, in general, to simple relationships between different asymmetric units in rotation space. The range of angles to be computed is a function of the symmetries of the individual Patterson function as discussed by Tollin, Main & Rossmann (1966).

Step 6. Matrices [ϱ] and hence [C] are calculated from the current values of the rotation angles.

Step 7. The Miller indices defined by p are generated systematically within a parallelepiped that completely envelops a hemisphere of reciprocal space to the given limits of $(\sin\theta/\lambda)$. Corresponding $|F_p|^2$ values are abstracted from the array in the core store. In the case of the 'large-term program' values of p and $|F_p|^2$ are
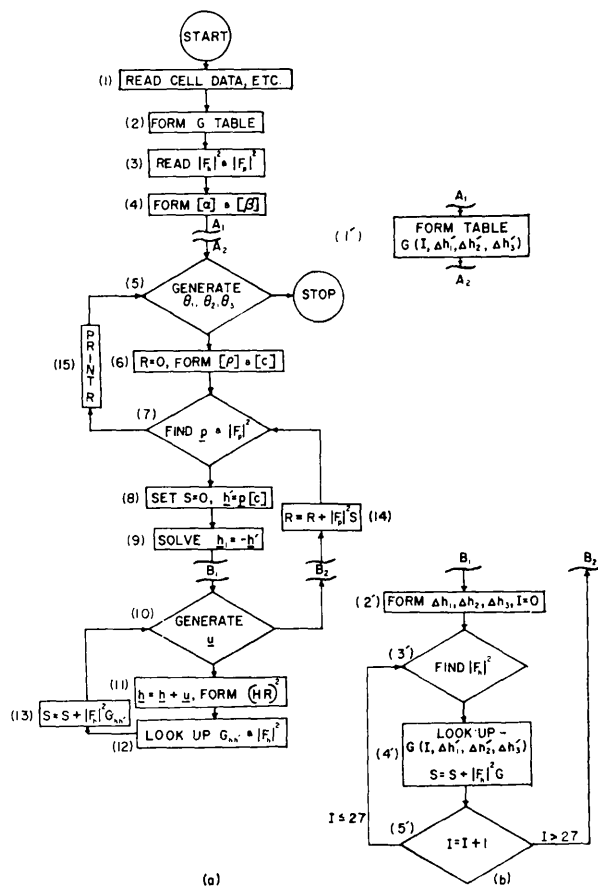

Fig. 2. Flow diagram of (a) the general program and (b) modification for the fine mesh program. For the latter program step 1' is inserted between the points A_1 and A_2, and the inner loops between B_1 and B_2 are replaced by steps 2' through 5'.
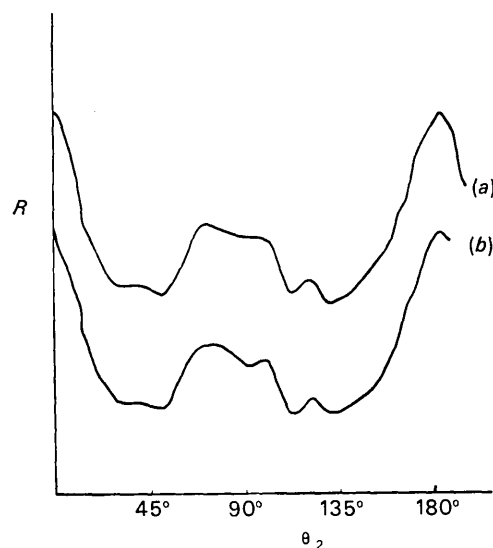

Fig. 3. Comparison of the monoclinic insulin rotation function results for (a) the large term program and (b) the fine mesh program.

not generated but are found by running sequentially through the list prepared in step 3.

Step 8. $\mathbf{h}'$ is evaluated and S, which will accumulate the sum $\{\sum_{\mathbf{h}} |F_{\mathbf{h}}|^2 G_{\mathbf{h},\mathbf{h}'}\}$, is set to zero.

Step 9. $\mathbf{h}_1$, the integral reciprocal lattice point in the first crystal nearest to the non-integral reciprocal lattice point $-\mathbf{h}'$ is found.

Steps 10 *and* 11. The n nearest neighbors around the point $\mathbf{h}_1$ are generated successively (this includes the point $\mathbf{h}_1$ itself). For the 27 nearest neighbors approximation, these points are given by

$$\mathbf{h}=\mathbf{h}_1+\mathbf{u},$$

where the components of u take on the values -1, 0, $+1$. $|\mathbf{H}|^2$, the square of the distance of the point $(\mathbf{h}+\mathbf{h}')$ from the origin of reciprocal space in the 'first' crystal can now be calculated.

Step 12. $|F_{\mathbf{h}}|^2$ is found in the data table for each value of $\mathbf{h}$. $G_{\mathbf{h},\mathbf{h}'}$ is found from the table constructed in step 2.

Step 13. The sum S is incremented by $|F_{\mathbf{h}}|^2 G_{\mathbf{h},\mathbf{h}'}$.

Step 14. R is incremented by $|F_{\mathbf{p}}|^2 S$, as soon as S has been evaluated for all the n nearest neighbors of $\mathbf{h}_1$ for the given $\mathbf{p}$.

Step 15. The values of the rotation angles, the corresponding magnitude of the rotation function R, and the elements of the matrix $[\mathbf{C}]$ are printed out.

The fine mesh program

The fine mesh program is a modified version of the general program. Step 1' is inserted between the points A_1 and A_2 (Fig. 2) in the flow diagram.

Step 1'. A table $G(I, \Delta h'_1, \Delta h'_2, \Delta h'_3)$ is constructed of the values of $G_{\mathbf{h},\mathbf{h}'}$ for each of the 27 points $\mathbf{h}$ with

respect to the mesh of non-integral points $\mathbf{h}'$. The latter is allowed to explore a three-dimensional grid $\Delta h'_1$, $\Delta h'_2$, $\Delta h'_3$ in steps of 0·2.

The innermost loops between the points B_1 and B_2 (Fig. 2) of the program are replaced by the part shown in Fig. 2(b). The steps involved are:

Step 2'. The components $\Delta h'_1$, $\Delta h'_2$, $\Delta h'_3$ are calculated.

Step 3'. h is determined from the current value of I, which is then used to find $|F_{\mathbf{h}}|^2$ from the array in the store.

Step 4'. The sum S is incremented by $|F_{\mathbf{h}}|^2 G_{\mathbf{h},\mathbf{h}'}(I, \Delta h'_1, \Delta h'_2, \Delta h'_3)$.

Step 5'. Steps 3' and 4' are repeated for all 27 values of I.

5. Strategy in the application of the rotation function

By using only the n nearest neighbors of $\mathbf{h}_i$ rather than all the vectors $\mathbf{h}$ as the inner sum, a cut-off has been applied to $G_{\mathbf{h},\mathbf{h}'}$. $G_{\mathbf{h},\mathbf{h}'}$ is the Fourier transform of a function defined by

$$F(r)=1 \quad \text{when } r \le r_0 \tag{3}$$

or

$$F(r)=0 \quad \text{otherwise,}$$

where $F(r)$ is defined in a space described by the spherical polar coordinates r, θ, φ. The Fourier transform of $F(r)$ is $G(H)$. $F(r)$ can then be expressed as

$$F(r)=\int_0^\infty G(H) \, \frac{\sin 2\pi rH}{2\pi rH} \cdot H^2 \cdot dH \tag{4}$$

(Patterson, 1959). The table G, evaluated by the program, was used to calculate the expression (4) between the limits 0 and H_0 (instead of infinity), in order to
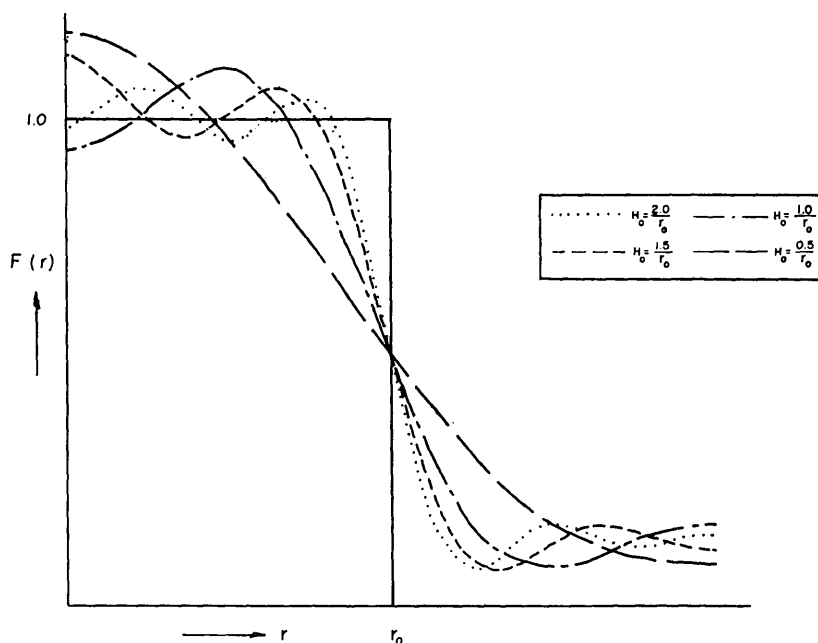

Fig. 4. The Fourier transform of the series representing $G(H)$ truncated at various values of H_0. The ideal shape of this curve is a step function indicated by the continuous lines.

test the effect of applying a cut-off to $G(H)$. The results for different H_0 values are compared with the ideal sphere of (3) in Fig. 4. It is clear that the main effect is to reduce the effective value of r_0. For values of $H_0 < 1/r_0$, $F(r)$ is a poor representation of a sphere. It is therefore important to consider carefully the choice of the number of terms, n, in the inner loop, as this number determines the effective value of H_0.

The effect of the termination of the series shows that it is important to select with care which crystal shall be P_1 and which P_2. The two criteria for this choice are rough equality of cell dimensions and lack of general systematic absences. If possible, a crystal that possesses both these properties should be chosen as P_1, or else the number, n, of integration points h must be increased in at least one direction.

A further consideration arises in case (b) of §1. It is helpful to place the known rigid group in an arbitrary unit cell of sufficient size to avoid overlap of the self-Patterson vectors, that is leaving gaps between the groups equal to the diameter of the group. An even larger cell leads only to increase in computing time without increase in clarity of the rotation function (Burnett, Tollin & Rossmann, to be published). The rigid group should also be placed in the unit cell in such a way that its symmetry, if any, becomes a space group operation. By so doing the Patterson function P_1 will have increased symmetry which will allow a reduction in the time for the computations. If the symmetry element is not so chosen then a non-linear symmetry operation in the rotation function results and the interpretation of the rotation function becomes cumbersome (Tollin, Main & Rossmann, 1966).

APPENDIX
Errata in Rossmann & Blow (1962)

(a) The expression for $\Omega_{hh'}$ in (5) should be

$$-\pi \sum_{i=1}^{3} (h_i + h'_i)(A_i^- + A_i^+) \,.$$

(b) The matrix $[\mathbf{C}]$ of equations (7) should be transposed.

(c) The footnote to page 26 is incorrect – the *same* convention was used by Patterson and hence

(d) The matrix $[\varrho]$ in Table 1(b) should be transposed.

(e) Equations (11) define only the magnitudes of $\kappa\psi\varphi$. The quadrant in which they lie must be obtained by comparing corresponding entries in the two terms of $[\varrho]$.

(f) The expression defining $\cos \omega$ following equation (9) should read as $\cos \omega = (\cos \alpha_2 - \cos \alpha_1 \cos \alpha_3)/\sin \alpha_1 \sin \alpha_3$.

(g) The minimum permissible angular interval discussed in §6 can be more usefully considered in terms of the resolution of the reflection data. If only data with spacing greater than d_{min} are included in the calculations and if the radius of the sphere of Patterson comparison is r_0, then a reasonable angular interval is $\frac{1}{2}d_{min}/r_0$. On this basis the Patterson function is moved through an angle which takes a point on the sphere through a distance equal to $\frac{1}{2}d_{min}$.

References

PATTERSON, A. L. (1959). *International Tables for X-ray Crystallography.* Vol. II, p. 73. Birmingham: Kynoch Press.

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

TOLLIN, P., MAIN, P. & ROSSMANN, M. G. (1966). *Acta Cryst.* **20**, 404.

TOLLIN, P. & SCOULOUDI, H. (1966). To be published.