

transitions will not yield a difference in the DT and hence only transitions involving a *klassengleich* or mixed-type symmetry reduction have to be considered. The allowed transitions between the different possible DT's are derived from the ABX_4 family tree and they are summarized in Fig. 6 where the symmetrically related typologies are enclosed between brackets.

Concluding remarks

For some of the transition sequences analysed in the present work, the Landau theory with the tilting of regular octahedra as the only order parameter occurring predicts the observed second-order character. This supports the hypothesis that in most of these transitions no other major structural distortions are present.

In those cases where a first-order character is found, either supplementary symmetry-reducing distortions or another possibility such as the lack of long-range order between octahedral layers may have to be considered. The presence of diffuse X-ray scattering in $RbAlF_4$ (Bulou *et al.*, 1983) supports the latter assumption.

For those transitions which were not mentioned in the discussion the second-order character has not yet been observed experimentally.

Finally, it may be concluded that the mere existence of a group-to-subgroup relation, according to the Birman-Worlock theorem, between ABX_4 phases which differ only in their tilt schemes is a sufficient condition for a second-order transition to be allowed.

I thank Professor Dr A. Janner and Dr E. Govaerts for valuable discussions. I am very indebted to Professor Dr S. A. Amelinckx and Professor Dr J. Van

Landuyt for their appreciated suggestions and the critical reading of the manuscript.

References

- ABRAHAMS, S. C. & BERNSTEIN, J. L. (1972). *Mater. Res. Bull.* **7**, 715-720.
- AHTEE, M., GLAZER, A. M. & MEGAW, H. D. (1972). *Philos. Mag.* **26**, 995-1014.
- BABEL, D., WALL, F. & HEGER, C. (1974). *Z. Naturforsch. Teil B*, **29**, 139-148.
- BRADLEY, C. J. & CRACKNELL, A. P. (1972). *The Mathematical Theory of Symmetry in Solids*. Oxford: Clarendon Press.
- BULO, A., FOURQUET, J. L., LEBLE, A., NOUET, J., DE PAPE, R. & PLET, F. (1982). *Solid State Chemistry, Proc. 2nd Eur. Conf.*, edited by R. METSELAAR, H. J. HEILIGERS & J. SCHOONMAN, pp. 679-682. Amsterdam: Elsevier Scientific.
- BULO, A., ROUSSEAU, M., NOUET, J., LOYZANCE, P. L., MOKHLISSE, R. & COUZI, M. (1983). *J. Phys. C*, **16**, 4527-4537.
- CRACKNELL, A. P. & JOSHUA, S. J. (1968). *J. Phys. A*, **1**, 40-42.
- DEBLIECK, R., VAN TENDELOO, G., VAN LANDUYT, J. & AMELINCKX, S. (1985). *Acta Cryst.* **B41**, 319-329.
- GLAZER, A. M. (1972). *Acta Cryst.* **B28**, 3384-3392.
- GLAZER, A. M. (1975). *Acta Cryst.* **A31**, 756-762.
- HERRING, C. (1942). *J. Franklin Inst.* **233**, 525-543.
- HIDAKA, M., FUJII, H., GARRARD, B. J. & WANKLYN, B. M. (1984). *Phys. Status Solidi A*, **86**, 75-83.
- HIDAKA, M., INOUE, K., GARRARD, B. J. & WANKLYN, B. M. (1982). *Phys. Status Solidi A*, **72**, 809-816.
- HIDAKA, M., WOOD, I. G. & GARRARD, B. J. (1979). *Phys. Status Solidi A*, **56**, 349-354.
- HIDAKA, M., WOOD, I. G., WANKLYN, B. M. & GARRARD, B. J. (1979). *J. Phys. C*, **12**, 1799-1807.
- International Tables for Crystallography* (1983). Vol. A. Dordrecht: Reidel.
- KAY, M. F. & BAILEY, P. C. (1957). *Acta Cryst.* **10**, 219-226.
- LANDAU, L. D. (1937). *Phys. Z. Sowjet Union*, **11**, 26.
- LIFSHITZ, E. M. (1941). *Zh. Eksp. Teor. Fiz.* **11**, 255-269.
- LOYZANCE, P. L. & COUZI, M. (1984). *Phys. Status Solidi A*, **85**, 359-374.
- LYUBARSKII, C. YA. (1960). *The Application of Group Theory in Physics*. Oxford: Pergamon Press.
- TOLEDANO, P. & TOLEDANO, J. C. (1982). *Phys. Rev. B*, **25**, 1946-1964.

Acta Cryst. (1986). **A42**, 325-334

Comparison of Vector Search and Feedback Methods for Finding Heavy-Atom Sites in Isomorphous Derivatives

BY MICHAEL G. ROSSMANN, EDWARD ARNOLD AND GERRIT VRIEND

Department of Biological Sciences, Purdue University, West Lafayette, Indiana 47907, USA

(Received 14 October 1985; accepted 25 February 1986)

Abstract

Heavy-atom sites in an isomorphous replacement derivative are usually found by inspection of a difference Patterson map. A systematic search procedure is necessary in the presence of high non-

crystallographic symmetry as in a crystalline virus. A reciprocal-space equivalent of the Patterson search procedure has been developed. Furthermore, it is shown that the Patterson search is closely analogous to the usual 'feedback' tests applied in checking a proposed site. The separation of self and cross vectors

in the Patterson map is shown to have its equivalent in reciprocal space. The procedure was used to identify the heavy-atom sites and the particle position in crystals of a human common cold virus.

Introduction

The classical method for solving macromolecular structure is the isomorphous replacement method (Green, Ingram & Perutz, 1954). This requires the initial determination of heavy-atom positions in an isomorphous derivative, usually by means of Patterson methods (Rossmann, 1960). Once the heavy-atom problem has been solved for the first heavy-atom compound, then single-isomorphous-replacement (SIR) phases can be calculated (Blow & Rossmann, 1961) and used to find lesser sites within the same compound or to solve the heavy-atom disposition in other derivatives. Generally, SIR phases are used to compute a difference electron density map to check heavy-atom sites by omitting known sites and phasing on tentative sites or *vice versa* (Dickerson, Kopka, Varnum & Weinzierl, 1967). This paper examines the relationship between the evidence for heavy-atom sites on the Patterson map and from feedback experiments.

The isomorphous replacement method can often be aided by the presence of non-crystallographic symmetry both in the determination of heavy-atom positions (Argos & Rossmann, 1976) and for the improvement of phases (*cf.* Rossmann & Blow, 1963; Argos, Ford & Rossmann, 1975; Bricogne, 1974). This enhancement of the isomorphous replacement method is essential in the structure determination of spherical viruses with icosahedral 532 symmetry.

The ideas presented in this paper arose in the analysis of an Au(CN)₂ derivative of human rhino (common cold) virus strain HRV14 (Rossmann, Arnold, Erickson, Frankenberger, Griffith, Hecht, Johnson, Kamer, Luo, Mosser, Rueckert, Sherry & Vriend, 1985). The crystals are cubic P2₁3 with $a = 445.1 \text{ \AA}$ and four virus particles per cell (Arnold, Erickson, Fout, Frankenberger, Hecht, Luo, Rossmann & Rueckert, 1984). One-third of the virus forms the asymmetric unit; thus there are 20 icosahedral asymmetric units per crystallographic asymmetric unit. The molecular weight of the virus is 8.5×10^6 or 142 000 dalton per icosahedral asymmetric unit. However, the RNA (33% by weight) is crystallographically disordered, leaving about 94 000 dalton of protein structure in which to find a few Au sites.

The Patterson search method

The usual Patterson procedure in the solution of heavy atoms in a spherical-virus derivative (Argos & Rossmann, 1976) consists of a systematic vector search. The icosahedral asymmetric unit is divided

into a grid with intervals less than one-third of the resolution limit of the data used in the Patterson calculation. The possibility of a heavy atom being in the position of any given grid point is then assessed by computing the equivalent positions in all non-crystallographic asymmetric units within the unit cell and, hence, the position of the corresponding point vectors. The latter are compared with the actual Patterson and the test criterion (say the sum of all sampled Patterson densities) is plotted on the corresponding grid point within the non-crystallographic asymmetric unit. A derivation of this procedure now follows.

Let the real difference Patterson density be $P_1(\mathbf{x})$ at the point \mathbf{x} and let the test Patterson density (derived from point atoms) be $P_2(\mathbf{x})$ at the point \mathbf{x} . Let the coefficients of the real difference Patterson be Δ_h^2 and the coefficients of the test Patterson be E_h^2 . A criterion of the degree of overlap, C , based on the sum of the Patterson densities at all test vectors would be

$$C = \int_V P_1(\mathbf{x})P_2(\mathbf{x}) \, d\mathbf{x},$$

where V is the volume of the unit cell. Therefore,

$$C = \sum_h \Delta_h^2 E_h^2. \quad (1)$$

Now the coefficients E_h can be computed from the presumed heavy-atom sites. When the presumed heavy-atom position on a grid point within the non-crystallographic asymmetric unit is multiplied by the non-crystallographic and crystallographic symmetry operators, it produces N equal sites within the unit cell. Let the i th site contribute a_i and b_i toward the real and imaginary parts of E_h and assume a relative scattering power of unity. Thus,

$$\begin{aligned} E_h^2 &= \left(\sum_N a_i \right)^2 + \left(\sum_N b_i \right)^2 \\ &= N + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^N (a_i a_j + b_i b_j). \end{aligned} \quad (2)$$

Therefore, from (1) and (2),

$$C = \sum_h \Delta_h^2 \left[N + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^N (a_i a_j + b_i b_j) \right].$$

Now $\sum_h \Delta_h^2 N$ is constant, independent of the given test site. Hence, the criterion C can be re-defined as

$$C = \sum_h \Delta_h^2 \left[\sum_{i=1}^{N-1} \sum_{j=i+1}^N (a_i a_j + b_i b_j) \right]. \quad (3)$$

Thus, computation of C from (3), for each test site in turn, is equivalent to the usual Patterson search procedure based on sampling the Patterson map at point vectors.

The single isomorphous difference Fourier synthesis

The phase probability distribution for a structure factor of a single isomorphous derivative has a mirror plane along the structure-factor direction of the heavy-atom contribution ($fe^{i\varphi}$) (Fig. 1). Thus if

$$(a) \quad |F_N| > |F_H|, \quad \alpha = \varphi + \pi$$

$$(\Delta = |F_H| - |F_N| \text{ is negative}) \text{ (Fig. 1a)}$$

or

$$(b) \quad |F_H| > (F_N^2 + f^2)^{1/2}, \quad \alpha = \varphi$$

$$(\Delta = |F_H| - |F_N| \text{ is positive}) \text{ (Fig. 1b)},$$

where F_N and F_H are the native and heavy-atom-derivative structure factors and α is the phase of F_N . The corresponding Fourier terms in the SIR difference density will, therefore, be

$$(a) \quad -|\Delta| \cos(2\pi \mathbf{h} \cdot \mathbf{x} - \varphi - \pi)$$

and

$$(b) \quad +|\Delta| \cos(2\pi \mathbf{h} \cdot \mathbf{x} - \varphi).$$

That is, the term will be $|\Delta| \cos(2\pi \mathbf{h} \cdot \mathbf{x} - \varphi)$ for both cases. Thus, in general and when Friedel's law holds, the SIR difference Fourier synthesis can be represented as

$$\rho(\mathbf{x}) = \sum_h |\Delta_h| \cos(2\pi \mathbf{h} \cdot \mathbf{x} - \varphi_h), \quad (4)$$

which is independent of the sign of the differences, Δ . This is somewhat surprising as the SIR map is nominally a phased Fourier summation, not a Patterson. However, there are a few intermediate cases where

$$|F_N| < |F_H| \quad \text{and} \quad |F_H| < (F_N^2 + f^2)^{1/2},$$

which will have negative terms in the SIR difference

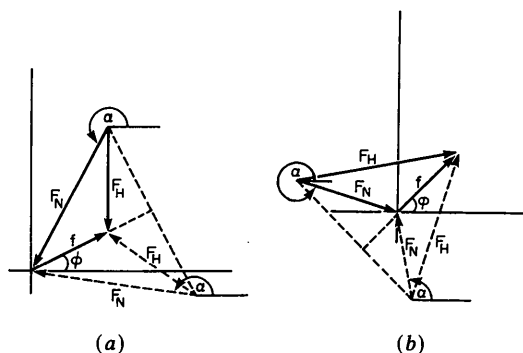


Fig. 1. Diagram for the phase solution of a SIR derivative. The two alternative solutions for the phase relationship $F_H = F_N + f$ are shown as a triangle with continuous lines and as a triangle with dashed lines. F_N , F_H and f are the structure factors for the native, heavy-atom derivative and heavy atoms alone. α is the phase of F_N and φ is the phase of f . (a) $|F_N| > |F_H|$ where $\alpha = \varphi + 180^\circ$. (b) $|F_N| < |F_H|$ where $\alpha = \varphi$.

Fourier synthesis. The sign of these terms will, however, be uncertain owing to inaccuracy in the relative scaling of the $|F_N|$, $|F_H|$ and $|f|$ amplitudes. That is, the sign of the Δ term is uncertain when $|\Delta|$ is small. However, as these terms will not contribute significantly to the total Fourier summation (4), this uncertainty is of little consequence. Nevertheless, it may be useful to multiply $|\Delta|$ by a figure of merit, m , which is dependent on the relative sizes of $|F_N|$, $|F_H|$ and $|f|$ (Appendix).

Feedback electron density

Let us suppose that a systematic Patterson search has been performed for all physically reasonable heavy-atom sites within the non-crystallographic asymmetric unit of an icosahedral virus crystal unit cell. However, the Patterson search function yields only a set of possible sites, rather than one or two outstanding sites. A way of testing the veracity of any one site is to use some of these sites for phasing a SIR difference Fourier synthesis and observing the appearance of other sites, hopefully consistent with the non-crystallographic symmetry. Alternatively, a specific site can be tested by omitting one of the J non-crystallographically related sites. The height of the difference Fourier synthesis can then be tested at the anticipated site of the omitted atom. This procedure can be done in J different ways. The criterion, S_j , for the quality of a proposed heavy-atom site at (x, y, z) can then be defined as

$$S(x, y, z) = \sum_{j=1}^J \rho_j, \quad (5)$$

where ρ_j is the density at atom j , an atom not used to compute the phase calculations. But from (4) (neglecting the few cases where the Fourier coefficients are small and negative),

$$\rho_j = \sum_h (|\Delta m|)_h \cos(2\pi \mathbf{h} \mathbf{x}_j - \varphi_{\neq j}), \quad (6)$$

where $\varphi_{\neq j}$ is the SIR phase of reflection \mathbf{h} computed on the basis of the J non-crystallographically related sites but omitting the atom at \mathbf{x}_j . Therefore from (5) and (6),

$$S(x, y, z) = \sum_h (|\Delta m|)_h \times \left[\sum_{j=1}^J (\cos 2\pi \mathbf{h} \mathbf{x}_j \cos \varphi_{\neq j} + \sin 2\pi \mathbf{h} \mathbf{x}_j \sin \varphi_{\neq j}) \right].$$

Now, if the real and imaginary components of the i th heavy-atom contribution are a_i and b_i , then

$$\cos \varphi_{\neq j} = (1/f_{\neq j}) \sum_{i \neq j} a_i \quad \text{and} \quad \sin \varphi_{\neq j} = (1/f_j) \sum_{i \neq j} b_i,$$

where

$$f_{\neq j} = \left[\left(\sum_{i \neq j} a_i \right)^2 + \left(\sum_{i \neq j} b_i \right)^2 \right]^{1/2}.$$

Also, by definition,

$$\cos 2\pi\mathbf{h}\mathbf{x}_j = a_j \quad \text{and} \quad \sin 2\pi\mathbf{h}\mathbf{x}_j = b_j.$$

With the assumption that $|f| = f_{\neq j}$ for all j , particularly where J is large, then

$$S(x, y, z) = \sum_h (|\Delta m/f|)_h \left[\sum_j \left(a_j \sum_{i \neq j} a_i + b_j \sum_{i \neq j} b_i \right) \right] \\ = 2 \sum_h (|\Delta m/f|)_h \left[\sum_{i=1}^{J-1} \sum_{j=i+1}^J (a_i a_j + b_i b_j) \right], \quad (7)$$

which is an expression very similar to (3) derived from considerations of Patterson searches. Thus, difference Patterson searches and feedback experiments based on SIR difference Fourier maps should give roughly equivalent results. The principal differences are:

(1) The Patterson search depends on a convolution of a point Patterson with a real Patterson. Representation of the point Patterson with a Fourier requires an infinite summation. However, only those terms for which there is an observed isomorphous difference can be included in the computation of C in (3).

(2) The effect of using coefficients Δ^2 in (3) or $|\Delta m/f|$ in (7). However, the behavior of m (Appendix) in (7) will, to a large extent, compensate for the squared expression in (3).

Computation of (7) will be rather unstable owing to the factor $|f|$ in the denominator of the Fourier coefficients as the coefficients for those terms with $|f|$ small might dominate the summation. In theory, this would be compensated by the necessary small $|\Delta|$ terms when $|f|$ is small. In practice, $|\Delta|$ may be larger due to experimental error in measuring small differences. This can be controlled by using the Fourier coefficients as Δm in (7) and setting m small whenever f is small (Appendix).

Searching for heavy-atom sites in the presence of known sites

Once one or more sites are known, the conventional procedure is to compute a SIR difference Fourier map, based on phasing from the known atoms, in order to ascertain whether there are other sites. This has its analogue in a feedback search procedure. However, a feedback procedure has the advantage of being able to use all but one of the non-crystallographically-related sites for additional phase information. From (4) we have

$$\rho(\mathbf{x}) = \sum_h |\Delta_h| \cos(2\pi\mathbf{h}\cdot\mathbf{x} - \varphi_h) \\ = \sum_h |\Delta_h| (\cos 2\pi\mathbf{h}\cdot\mathbf{x} \cos \varphi_h + \sin 2\pi\mathbf{h}\cdot\mathbf{x} \sin \varphi_h).$$

Now let A_h and B_h be the real and imaginary components of structure factors due to the known atoms alone. Also assume equal occupancy for the known

sites and the current test site. Hence, the electron density at the j th non-crystallographically-related search site will be given by

$$\rho(\mathbf{x}_j) = \sum_h |\Delta m/f|_h \left[\cos 2\pi\mathbf{h}\cdot\mathbf{x}_j \left(A + \sum_{i \neq j} a_i \right) + \sin 2\pi\mathbf{h}\cdot\mathbf{x}_j \left(B + \sum_{i \neq j} b_i \right) \right]$$

on weighting by a figure of merit and on computing phases based on the known atoms plus all but the j th non-crystallographically-related test sites. Therefore,

$$\rho(\mathbf{x}_j) = \sum_h |\Delta m/f|_h \left[a_j \left(A + \sum_{i \neq j} a_i \right) + b_j \left(B + \sum_{i \neq j} b_i \right) \right].$$

If then the test criterion, S , is given by (5)

$$S(x, y, z) = \sum_j \sum_h |\Delta m/f|_h \left[a_j \left(A + \sum_{i \neq j} a_i \right) + b_j \left(B + \sum_{i \neq j} b_i \right) \right].$$

Now if

$$\sum_j a_j = a \quad \text{and} \quad \sum_j b_j = b,$$

it follows that

$$S(x, y, z) = \sum_h |\Delta m/f|_h \left[(Aa + Bb) + \sum_{i,j} (a_i a_j + b_i b_j) \right]. \quad (8)$$

This is a more general form of (7) that includes the additional $(Aa + Bb)$ term due to the known heavy-atom sites.

As before, the same expression can be readily approached from the point of view of a Patterson search procedure. The coefficients of the point Patterson synthesis will now be given by (assuming equal occupancy for the test and search sites)

$$E^2 = \left(A + \sum_i a_i \right)^2 + \left(B + \sum_i b_i \right)^2.$$

Hence, as before, using (1), the criterion C will be given by

$$C(x, y, z) = \sum_h \Delta_h^2 \left[(A^2 + B^2) + \sum_i (a_i^2 + b_i^2) + \left(A \sum_i a_i + B \sum_i b_i \right) + \sum_{i,j} (a_i a_j + b_i b_j) \right].$$

Again, as before, the quantities $(A^2 + B^2)$ and $(a_i^2 + b_i^2)$ will be constant and, therefore, independent of

the currently assumed test site. Hence, the test criterion can be written as

$$C(x, y, z) = \sum_h \Delta_h^2 \left[(Aa + Bb) + \sum_i \sum_j (a_i a_j + b_i b_j) \right], \quad (9)$$

which is closely similar to (8) derived from feedback considerations.

The advantage of the feedback computations in reciprocal space or the Patterson search method over the classical SIR difference electron density is that the phasing effect of the current test site can be included along with the phasing effect of the 'known' sites. Therefore, if the 'known' sites are themselves tentative, the joint phasing effect will have greater power than is exhibited in the SIR Fourier. In other words:

(1) the SIR Fourier series is equivalent to the computation of

$$P(x, y, z) = \sum_h \Delta_h^2 (Aa + Bb),$$

where A, B are the contributions of the known sites and a, b are the contributions of a test site at (x, y, z) and its symmetry-related positions;

(2) the Patterson function (or reciprocal-space feedback experiment), when based on only one test site, is based on the terms

$$Q(x, y, z) = \sum_h \Delta_h^2 \sum_i \sum_j (a_i a_j + b_i b_j),$$

where a_i, b_i are the contribution of a test site at the i th symmetry-equivalent position x, y, z ; and

(3) the Patterson function (or reciprocal-space feedback) is based on the sum

$$C(x, y, z) = P(x, y, z) + Q(x, y, z)$$

when one or more sites are already tentatively known.

Should one or more sites be already known or tentatively known, this also implies a solution of the translation problem of placing a particle (e.g. a virus) in the unit cell relative to the crystallographic symmetry axes.

The effect of crystallographic and non-crystallographic symmetry

Let there be I crystallographic asymmetric units in the unit cell and J non-crystallographic asymmetric units within one crystallographic asymmetric unit. Then let a_{ij} be the real component of a given heavy atom in the j th non-crystallographic asymmetric unit of the i th crystallographic asymmetric unit. Now the feedback method depends on omitting the $j = k$ th atom in every crystallographic asymmetric unit and computing the phase, $\varphi_{\neq j}$, based on all the other $J - 1$

sites. Hence,

$$\begin{aligned} & f_{\neq j} \cos 2\pi \mathbf{h} \cdot \mathbf{x}_j \cos \varphi_{\neq j} \\ &= \left[a_{1k} \left(\sum_{j \neq k} a_{1j} + \sum_{j \neq k} a_{2j} + \dots \right) \right. \\ & \quad \left. + a_{2k} \left(\sum_{j \neq k} a_{1j} + \sum_{j \neq k} a_{2j} + \dots \right) + \dots \right] \\ &= \sum_s \sum_t A_{st}, \end{aligned}$$

where

$$A_{st} = \sum_{\substack{p=1 \\ p \neq q}}^J \sum_{q=1}^J a_{sp} a_{tq}.$$

A similar expression can be derived for the imaginary components. That is, all terms along diagonals of squares representing interactions between atoms in crystallographic units s and t are omitted (Fig. 2). These interactions are the Harker peaks in Patterson space. The Patterson search method thus differs also from the feedback method by including the Harker interactions.

A further difference between the feedback and Patterson methods occurs near non-crystallographic rotation symmetry axes. If a trial heavy atom is placed near an n -fold rotation axis, then it will generate $n - 1$ other atoms near the trial position. If one of these is omitted to test the effect of feedback, then the feedback test site will automatically contain a large peak as it overlaps the assumed heavy-atom site. In a Patterson search procedure this situation is avoided by omitting the short vectors near the origin. Similar exclusions could be used in the feedback method in order to avoid the spurious large peaks on rotation axes.

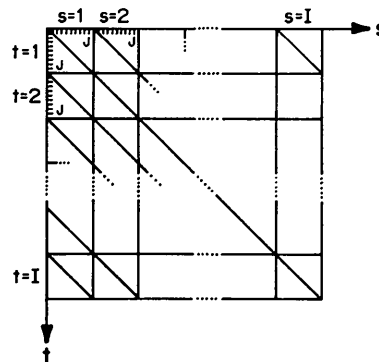


Fig. 2. Interactions used for feedback and Patterson calculations. There are I crystallographic units and J non-crystallographic units within each of these. All interactions along the diagonal $s = t, s + 1 = t, s + 2 = t$ etc. are omitted in the calculations of the feedback criterion (7). In contrast, a Patterson calculation omits only the terms along $s = t$.

Factorization to separate rotational and translational parameters

The criteria (3) or (7) must be repeatedly calculated for successive trial positions of a heavy atom. Let us take, for instance, a virus particle. The trial position can be confined to a search volume defining the non-crystallographic asymmetric unit as well as between physically reasonable external and internal radial limits representing the probable thickness of the coat protein. However, the particle position needs also to be defined. In many cases, such as for the solution of the southern bean mosaic virus structure (Rayment, Johnson, Suck, Akimoto & Rossmann, 1978) or the cowpea mosaic virus structure (Johnson, private communication), the particle center is fixed by a special position. In other cases, such as the rhinovirus structure determination (Arnold *et al.*, 1984), the virus may sit on a rotation axis giving one degree of freedom to its motion. In yet other cases, the virus position is unconstrained by crystallographic symmetry [e.g. satellite tobacco necrosis virus (Lentz, Strandberg, Unge, Vaara, Borell, Fridborg & Petef, 1976)]. In general, once the particle orientation is known there exists a six-dimensional problem when locating a heavy-atom site from difference data: three heavy-atom parameters and three particle (molecular) parameters. A systematic search procedure would thus be prohibitively expensive unless (as is so often the case) the position of the particle center is known or severely limited. However, when the search problem is couched in reciprocal space, as in (3), (7), (8) and (9) above, it is possible to factor out the translational components.

Let the j th (purely rotational) non-crystallographic symmetry operation be represented by

$$\mathbf{x}_j = [C_j]\mathbf{x},$$

where \mathbf{x} is a position vector expressed as fractional cell lengths with respect to the particle center. Then, if \mathbf{u} is the positional vector of the center of the 'standard' particle in fractional coordinates, the i th crystallographic symmetry operator can be represented as

$$\mathbf{x}_{ij} = [T_i](\mathbf{x}_j + \mathbf{u}) + \mathbf{d}_i$$

or

$$\mathbf{x}_m = [T_i]\mathbf{u} + [T_i][C_j]\mathbf{x} + \mathbf{d}_i,$$

where m is the j th non-crystallographic position in the i th crystallographic asymmetric unit. Then the expression which needs to be evaluated in (3) or (7) can be written as

$$\begin{aligned} a_m a_n + b_m b_n &= \cos(2\pi \mathbf{h} \cdot \mathbf{x}_n) \cos(2\pi \mathbf{h} \cdot \mathbf{x}_m) \\ &+ \sin(2\pi \mathbf{h} \cdot \mathbf{x}_n) \sin(2\pi \mathbf{h} \cdot \mathbf{x}_m) \\ &= \cos[2\pi \mathbf{h}(\mathbf{x}_m - \mathbf{x}_n)] \\ &= \cos\{2\pi \mathbf{h}([T_{i1}] - [T_{i2}])\mathbf{u} + ([T_{i1}][C_{j1}] \\ &- [T_{i2}][C_{j2}])\mathbf{x} + (\mathbf{d}_{i1} - \mathbf{d}_{i2})\}, \end{aligned}$$

where $i1$ refers to the crystallographic unit of the m th atom and $i2$ refers to the crystallographic unit of the n th atom. Similarly, $j1$ refers to the non-crystallographic asymmetric unit of the m th atom and $j2$ refers to the non-crystallographic asymmetric unit of the n th atom. Thus,

$$a_m a_n + b_m b_n = \cos(2\pi \mathbf{h}' \cdot \mathbf{u} + \theta_{h'}), \quad (10)$$

where

$$\mathbf{h}' = \mathbf{h}([T_{i1}] - [T_{i2}]); \quad (11)$$

$$\theta_{h'} = \mathbf{h}([T_{i1}][C_{j1}] - [T_{i2}][C_{j2}])\mathbf{x} + (\mathbf{d}_{i1} + \mathbf{d}_{j2}).$$

Now \mathbf{h}' is integral and can be evaluated from the known crystallographic operators. Similarly, $\theta_{h'}$ can be evaluated for a given position \mathbf{u} of the chosen heavy-atom site relative to the particle center. Hence, from (3) or (7), the criterion for a heavy-atom site at \mathbf{u} can be evaluated as, say,

$$S(\mathbf{u}) = 2 \sum_h (|\Delta m|)_h \left[\sum_i \sum_j \cos(2\pi \mathbf{h}' \cdot \mathbf{u} + \theta_{h'}) \right], \quad (12)$$

which is a Fourier summation that can be readily and quickly evaluated for all values of \mathbf{u} .

It is of some interest to examine the values of \mathbf{h}' from (11). When considering interactions between atoms related by non-crystallographic symmetry (different j) but in the same crystallographic asymmetric unit or particle (same i), then $[T_{i1}] = [T_{i2}]$ and $\mathbf{h}' = 0$. In this case, $\theta_{h'=0} = \mathbf{h}[T_{i1}][C_{j1}] - [C_{j2}]\mathbf{x}$. Thus, the summation (12) can be written as

$$\begin{aligned} S(\mathbf{u}) &= 2I \sum_h (|\Delta m|)_h \left[\sum_{\substack{j \\ \text{self}}} \cos \theta_{h'=0} \right] \\ &+ 2 \sum_h (|\Delta m|)_h \left[\sum_{\substack{i \\ \text{cross}}} \sum_j \cos 2\pi(\mathbf{h}'\mathbf{u} + \theta_{h'}) \right]. \end{aligned}$$

The self terms (between atoms in the crystallographic asymmetric unit) correspond to searching for the self vectors around the origin of a Patterson function. The cross terms (between crystallographic asymmetric units) are additional terms, which modify the self terms and correspond to the strengthening of a Patterson search by incorporating the cross vectors between particles. Hence, $S(\mathbf{u})$ can be computed without adding the cross terms in an attempt to determine the heavy-atom sites relative to the particle center when the position of the particle center is as yet unknown.

The effect of a twofold crystallographic axis, say along \mathbf{b} , would give values of \mathbf{h}' as

$$\begin{aligned} \mathbf{h}' &= \mathbf{h} \left[\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} - \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \right] \\ &= 2\mathbf{h}, 0, 2\mathbf{l}. \end{aligned}$$

Thus, the search function would be two dimensional since there is no significance to a translation parallel to b .

Computational search procedures

A program was written for the Cyber 205 supercomputer. The advantage of the supercomputer is the vectorization options whereby repetitive operations may be speeded greatly. This facility was exceedingly important in the structure-factor computation where

the quantity $\sum \sum_{i \neq j} (a_i a_j + b_i b_j)$ has to be computed for all atoms in the cell per search point [see (3), (7), (8) and (9) above]. For cubic rhinovirus crystals there are 240 symmetry (crystallographic and non-crystallographic)-equivalent points per cell.

The flow chart of the program is shown in Fig. 3. The cell dimensions are used for the orthogonalization and de-orthogonalization matrices $[\alpha]$ and $[\beta]$ (in the routine *RECIP*) as defined by Rossmann & Blow (1962). Here

$$\mathbf{x} = [\alpha]\mathbf{X} \quad \text{and} \quad \mathbf{X} = [\beta]\mathbf{x},$$

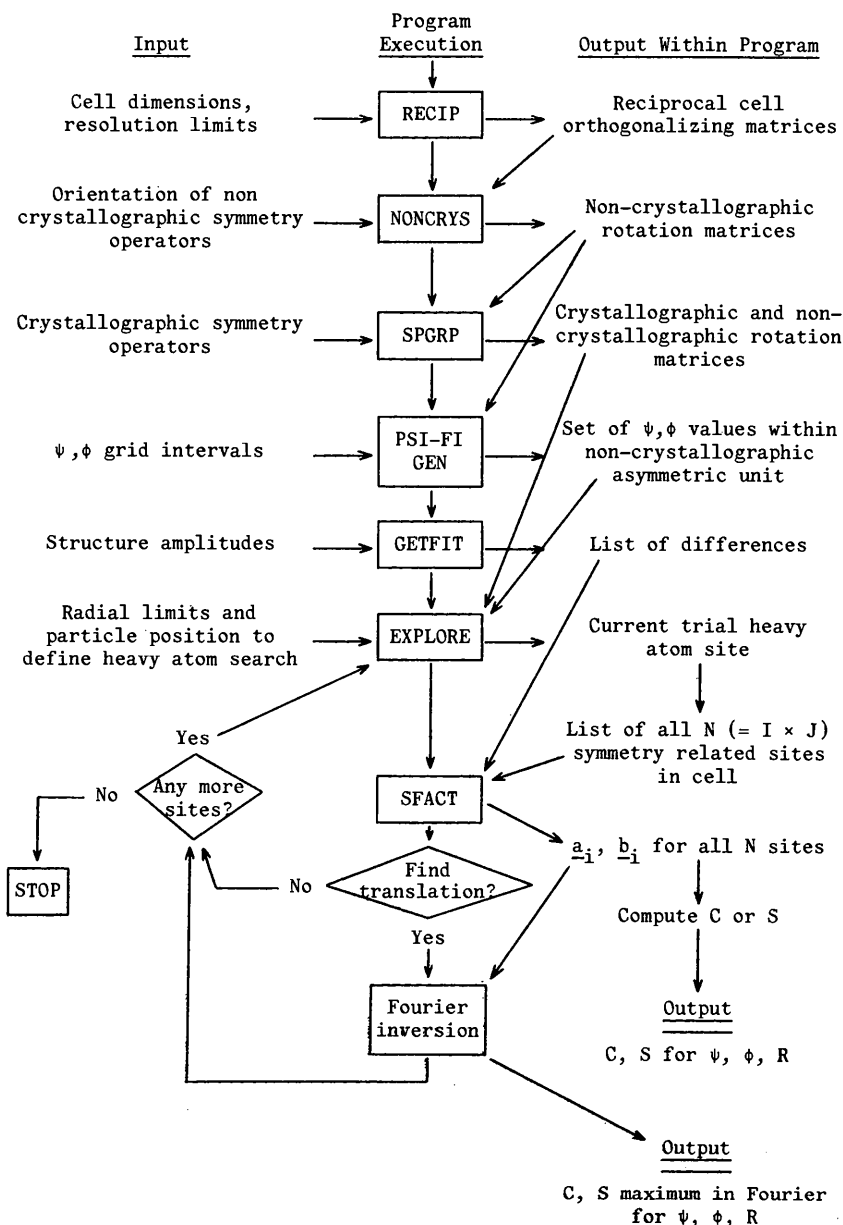


Fig. 3. Flow chart for heavy-atom search.

where \mathbf{X} is the position vector of a point within the particle with respect to a defined origin (like the particle center). The components of \mathbf{X} have dimensions of length, while \mathbf{x} are fractional coordinates of this point relative to the cell lengths.

The unique non-crystallographic operator properties are provided in terms of the ψ , φ polar angles which a given rotation axis makes together with the cell axes and the rotation, κ , about this axis. The orientations of these axes will have been determined perhaps from a rotation function. Thus, for a particle with 222 symmetry it will be necessary to read two operators, while for an icosahedral particle it will be necessary to read four operators. Each operator is used in the routine *NONCRYST* to set up the rotation matrix $[\rho]$ given ψ , φ and κ (Rossmann & Blow, 1962). This matrix is then used by the routine *NONCRYST* to operate in turn on itself (e.g. five times for $\kappa = 72^\circ$) and on the other $[\rho]$ matrices, producing a total of J non-crystallographic rotation matrices of the form

$$[C_j] = [\alpha][\rho_j],$$

where

$$\mathbf{x} = [\alpha][\rho_j]\mathbf{X}.$$

The I crystallographic operators ($[T_i]\mathbf{x} + \mathbf{d}_i$) can then be incorporated (in the routine *SPGRP*) to give $I \times J$ matrices of the form $[T_i][C_j]$.

It is then necessary to set up a grid, in polar coordinates ψ and φ , which explores all sites within the non-crystallographic asymmetric unit (see the shaded area in Fig. 3). This can be done by considering every position of a suitable grid in a hemisphere ($0 \leq \psi \leq \pi$, $0 \leq \varphi < 2\pi$). Each combination of ψ and φ is converted into the Cartesian coordinates $X = \cos \varphi \sin \psi$, $Y = \sin \varphi \sin \psi$, $Z = \cos \psi$. All non-crystallographically-related positions are then generated for each ψ , φ position using the matrices computed earlier by the routine *NONCRYST*. Only those positions of ψ and φ are accepted where the first (say) non-crystallographic operator also gives the smallest packed quantity $10^{10}(X+A) + 10^5(Y+B) + Z+C$, where A , B and C are suitable biasing constants. The resultant positions will represent one non-crystallographic asymmetric unit for the given orientation of the non-crystallographic operators.

A list of structure-factor differences is then selected from a suitable input tape by the routine *GETFIT*. This includes selection within a suitable resolution range, computation of mean differences for subsequent figure-of-merit calculations *etc.* The amplitudes and their indices are stored in arrays (vectors) for the structure-factor routine.

The routine *EXPLORE* steps through each ψ , φ position systematically between selected radial limits, R , of the heavy-atom position relative to the particle center. With the selected position and the earlier

evaluated matrices, all $I \times J$ sites in the unit cell can then be calculated.

The major calculational effort occurs in the routine *SFACT* which evaluates the Fourier terms for (12) with respect to a given site (ψ , φ , R) and sums over all \mathbf{h} reflections. Should it be necessary to explore a translation then these coefficients are inverted by a Fourier routine before proceeding to the next proposed heavy-atom site.

The actual computational times using the reciprocal-space search procedure were longer than the corresponding Patterson search when applied to 6 Å resolution data for an Au(CN)₂ derivative of cubic rhinovirus crystals. Furthermore, given the present vectorization algorithm, the times would increase substantially with increasing resolution. Nevertheless, the reciprocal-space method gives far greater flexibility as the area of a peak can be covered many times using different resolution cutoffs, different forms for the coefficients and a variety of reflection rejection criteria. Thus, it is much faster to cover a few points with different conditions with the reciprocal-space method, but it is faster to search the whole non-crystallographic asymmetric unit with the Patterson-search technique.

A typical time for the cubic rhinovirus difference search was 2' per point using 22 354 reflections to 6 Å resolution. A tenfold improvement, concomitant with an improvement of the results, could be obtained by selecting only the large-term differences. The improved results were due to the omission of error implicit in the smaller differences.

Results

The difference between the Patterson (3) and feedback (7) criteria was tested in an application to cowpea mosaic virus (CpMV). The data were kindly supplied by Professor Jack Johnson of Purdue University. He and his colleagues have recently determined the heavy-atom site for an Hg derivative of CpMV crystals using the classical Patterson search method. The derivative turned out to have a single-site substitution of exceptional quality. Hence, the data were used for testing the procedure (Table 1). The crystals of CpMV are cubic *I*23 (White & Johnson, 1980) with $a = 315.0$ Å. There are two particles per cell with each particle situated on a 23 special position. Data were selected between 15.0 and 5.5 Å resolution using only reflections with $|F| \geq 2\sigma(F)$. There were then, in total, 8717 independent significant reflections. Table 1 shows the extraordinarily clean search result. However, there appears to be little difference whether the Δ^2 or $|\Delta m|$ coefficients were used in the procedure. Application to the Au(CN)₂ derivatives of the human rhinovirus 14 data will be given in detail in a subsequent publication (Arnold, Vriend, Luo, Griffith, Kamer, Erickson, Johnson &

Table 1. *Cowpea mosaic virus peak exploration using the feedback search method*

The peak values, normalized to a maximum value of 100, are shown for the search positions given by the ψ , φ and R values. The terms used for evaluating the search function are described in the text.

(a) Δ^2 coefficients

	$R = 100 \text{ \AA}$			$R = 110 \text{ \AA}$			$R = 120 \text{ \AA}$		
$\varphi \rightarrow$	-78	-77	-76	-78	-77	-76	-78	-77	-76
$\psi = 84$	7	14	14	9	41	39	11	25	19
85	13	16	11	32	100	65	5	6	4
86	11	12	0	31	62	34	9	6	4

(b) $|\Delta m|$ coefficients

	$R = 100 \text{ \AA}$			$R = 110 \text{ \AA}$			$R = 120 \text{ \AA}$		
$\varphi \rightarrow$	-78	-77	-76	-78	-77	-76	-78	-77	-76
$\psi = 84$	13	10	17	9	39	39	14	26	20
85	13	15	16	32	100	63	7	3	16
86	11	10	0	29	53	34	9	7	6

Table 2. *Effects of differences of various sizes*

Contributions to the peak at $\psi = 85^\circ$, $\varphi = -77^\circ$, $R = 110 \text{ \AA}$ in the CpMV heavy-atom search (Table 1a) are divided into components dependent on the size of the differences.

Size of difference	n	Contribution to sum
0.0-0.25	1484	-0.1
0.25-0.50	1418	-0.6
0.50-1.00	2260	-2.2
1.00-2.00	2606	17.5
2.00 and larger	945	85.4
Total	8717	100.0

Note: Size differences are given as a fraction of the r.m.s. difference for all the observed reflections. n is the number of reflections in each range.

Rossmann, in preparation) relating to the complete structure determination of that virus (Rossmann *et al.*, 1985).

Tests were conducted on the effect of excluding the small differences from the calculations (Table 2). The initial tests were conducted on the heavy-atom peak shown for CpMV in Table 1. The differences were divided into ranges as a fraction of their size relative to the r.m.s. difference, $\langle \Delta \rangle$, of all observed reflections. It is clear that only the reflections greater than $1\langle \Delta \rangle$ made useful contributions to the peak. Although only the center of the peak is shown in Table 2, the same observations were valid for points off the central peak position. Hence at best only 40% of the reflections gave useful information, whereas the other 60% of the data actually created interference. Similar observations were apparent for the Au(CN)₂ sites in human rhinovirus 14. In this case, however, only the 16% largest differences contributed to the peak. Thus by omitting 80 to 90% of the data, in a manner analogous to the 'big-term' rotation function (Tollin & Rossmann, 1966), the speed of the program could be vastly improved, while at the same time the results were cleaner.

Concluding remarks

The Patterson search procedure for finding heavy-atom sites for an isomorphous heavy-atom derivative is equivalent to using a feedback procedure based on phases from presumed sites and testing for the appearance of a non-crystallographically-related site. The use of self and cross vectors in the Patterson search method has its analogy in reciprocal space. The procedure can be carried out in reciprocal space with greater convenience of the precise coefficients and resolution ranges to be tested, yielding potentially more accurate results. However, the feedback search method is several hundred times slower than the Patterson search method given the present computer algorithms. A reasonable strategy is to perform an exhaustive Patterson search to locate promising heavy-atom sites and then to explore these using different conditions with the feedback search method.

We are grateful to Dr Jack Johnson for permitting us to use his CpMV data to test the reciprocal-space search program. The work was supported by grants from the National Institutes of Health and the National Science Foundation to MGR and a Damon Runyon-Walter Winchell postdoctoral fellowship to EA.

APPENDIX

Computation of figures of merit

Let us assume that $|F_N| < |F_H|$ (Fig. 1a). Then lack of closure when $\alpha = \varphi + \pi$ is

$$\begin{aligned} \varepsilon_1 &= |(F_N + kf)| - |F_H| \\ &\approx k|f| - |\Delta|, \end{aligned}$$

where k is the scale factor to place the calculated heavy-atom structure factor $|f|$ onto the same relative scale as the observed differences $|\Delta|$ and $\Delta = |F_H| - |F_N|$. Thus, the phase probability for the phase (φ) is given by

$$P_1(\varphi) \propto \exp[-(\varepsilon_1^2/2E^2)],$$

where E is the standard error for the lack of closure and can be calculated as $(\sum_h \Delta^2/h)^{1/2}$. Similarly, when $\alpha = \varphi$

$$\begin{aligned} \varepsilon_2 &= |(F_N - kf)| - |F_H| \\ &\approx -k|f| - |\Delta| \end{aligned}$$

and

$$P_2(\varphi + \pi) \propto \exp[-(\varepsilon_2^2/2E^2)].$$

Similar expressions can be derived for the case shown in Fig. 1(b).

Hence, if all phase probabilities other than those at φ and $\varphi + \pi$ are neglected,

$$m = (P_1 - P_2)/(P_1 + P_2).$$

The quantities P_1 and P_2 can be computed provided there exists a knowledge of the scale factor k . This can be estimated by assuming

$$k\langle f \rangle = 2^{1/2}\langle \Delta \rangle,$$

where $\langle \rangle$ signifies mean quantities. Now $\langle f \rangle = J^{1/2}$ if there are J atoms of unit weight in the cell. Hence,

$$k = \langle \Delta \rangle (2/J)^{1/2}.$$

It follows that if, for a particular reflection, $f = 0$, then $P_1 = P_2$ and $m = 0$. Similarly, if $|\Delta| = 0$, then $P_1 = P_2$ and $m = 0$. Indeed, the expression for m is entirely symmetrical between $k|f|$ and $|\Delta|$. The figure of merit is, therefore, small whenever $k|f|$ or $|\Delta|$ are small compared to $\langle \Delta \rangle$.

Acta Cryst. (1986). **A42**, 334-339

Distributions of Sums and of Ratios of Sums of Intensities

BY A. J. C. WILSON

*Crystallographic Data Centre, University Chemical Laboratory,
Lensfield Road, Cambridge CB2 1EW, England*

(Received 14 February 1985; accepted 7 March 1986)

Abstract

If intensities of reflexion have one or other of the ideal distributions [Wilson (1949). *Acta Cryst.* **2**, 318-320], the sum of n such intensities has a gamma (γ) distribution with parameter n (acentric) or $n/2$ (centric), and the ratio of two such sums has a beta (β) distribution. These distributions are applied to (i) intensities normalized to the ideal average Σ ; (ii) intensities normalized to the local average $\langle I \rangle$; (iii) ratios used for scaling. Bias in scaling is discussed, and certain results are obtained for non-ideal distributions. Expressions are obtained for the variance of the traditional reliability index R for both ideal distributions; these have applications in certain methods of structure determination [Rabinovich & Shakked (1984). *Acta Cryst.* **A40**, 195-200].

0108-7673/86/050334-06\$01.50

References

- ARGOS, P., FORD, G. C. & ROSSMANN, M. G. (1975). *Acta Cryst.* **A31**, 499-506.
 ARGOS, P. & ROSSMANN, M. G. (1976). *Acta Cryst.* **B32**, 2975-2979.
 ARNOLD, E., ERICKSON, J. W., FOUT, G. S., FRANKENBERGER, E. A., HECHT, H. J., LUO, M., ROSSMANN, M. G. & RUECKERT, R. R. (1984). *J. Mol. Biol.* **177**, 417-430.
 BLOW, D. M. & ROSSMANN, M. G. (1961). *Acta Cryst.* **14**, 1195-1202.
 BRICOGNE, G. (1974). *Acta Cryst.* **A30**, 395-405.
 DICKERSON, R. E., KOPKA, M. L., VARNUM, J. C. & WEINZIERL, J. E. (1967). *Acta Cryst.* **23**, 511-522.
 GREEN, D. W., INGRAM, V. M. & PERUTZ, M. F. (1954). *Proc. R. Soc. London Ser. A*, **225**, 287-307.
 LENTZ, P. J. JR, STRANDBERG, B., UNGE, T., VAARA, I., BORELL, A., FRIDBORG, K. & PETEF, G. (1976). *Acta Cryst.* **B32**, 2979-2983.
 RAYMENT, I., JOHNSON, J. E., SUCK, D., AKIMOTO, T. & ROSSMANN, M. G. (1978). *Acta Cryst.* **B34**, 567-578.
 ROSSMANN, M. G. (1960). *Acta Cryst.* **13**, 221-226.
 ROSSMANN, M. G. & BLOW, D. M. (1962). *Acta Cryst.* **15**, 24-31.
 ROSSMANN, M. G. & BLOW, D. M. (1963). *Acta Cryst.* **16**, 39-45.
 ROSSMANN, M. G., ARNOLD, E., ERICKSON, J. W., FRANKENBERGER, E. A., GRIFFITH, J. P., HECHT, H. J., JOHNSON, J. E., KAMER, G., LUO, M., MOSSER, A. G., RUECKERT, R. R., SHERRY, B. & VRIEND, G. (1985). *Nature (London)*, **317**, 145-153.
 TOLLIN, P. & ROSSMANN, M. G. (1966). *Acta Cryst.* **21**, 872-876.
 WHITE, J. M. & JOHNSON, J. E. (1980). *Virology*, **101**, 319-324.

1. Introduction

1.1. Notation

In several crystallographic contexts it is necessary to consider sums like

$$J_n = \sum_{i=1}^n G_i, \quad (1)$$

$$K_m = \sum_{i=1}^m H_i, \quad (2)$$

and ratios like

$$S_{n,m} = J_n / K_m, \quad (3)$$

where G_i and H_i are the intensities of sets of reflexions. Similar expressions where G_i and H_i are