size $t_h$ calculated from (6),

$$t_h = 0.94\lambda/\{\cos\theta_h[\Delta 2\theta(\text{FWHM}) - \delta - 2\Delta\lambda/\lambda \tan\theta_h]\}, \quad (7)$$

is given in the fourth column. Fig. 4, as well as the results given in Table 3, agrees well with the TEM result, confirming the applicability of the approach introduced in this paper.

## Discussion

Equation (2) is applicable only to ideal powder samples, *i.e.* for powders consisting of strain-free and preferred-orientation-free crystallites, whose orientations are completely random and the number of which present in the irradiated specimen is very large. Furthermore, it was assumed that extinction and absorption effects can be neglected. Only such an ideal powder is represented in reciprocal space by ideally spherical concentric shells with homogenous thickness and occupation density.

Strain, preferred orientation, large absorption and/or extinction result in deformations of these shells, varying $\mathbf{d}^*$ (strain), $\varepsilon$ (absorption, extinction, strain) and the occupation density within the shells (preferred orientation). It will be shown in paper II that all these effects can be represented in reciprocal space and have to be taken properly into account, resulting in general equations analogous to (2).

## Concluding remarks

Comparing the width $\Delta\theta_h$ obtained for single-crystal diffractometry, expression (1*b*), with the formula for $\Delta 2\theta$, (6), deduced for powder diffraction, it is obvious that the main difference is introduced by the term corresponding to the size of the coherently scattering particles. The random orientation of the crystallites in the powder results in $\cos\theta_h$ in the denominator of this term, causing appreciable broadening of the Bragg reflections for large Bragg angles only. In single-crystal diffractometry, the crystal is rotated about the $\theta$ axis during the scan. This rotation causes $\sin 2\theta_h$ in the denominator of the particle-size term in (1*b*). The rocking curves are therefore broadened appreciably for small as well as for large values of $\theta_h$.

I thank Mrs G. Kumpat and Mr G. Ulrich for preparing the computer graphics of Figs. 1–4.

### References

AZAROFF, L. V. (1968). *Elements of X-ray Crystallography*. New York: McGraw-Hill.
BOLIS, V., FUBINI, B., GIAMELLO, E. & RELLER, A. (1989). *J. Chem. Soc. Faraday Trans.* 1, **85**, 855–867.
CAGLIOTI, G., PAOLETTI, A. & RICCI, F. P. (1958). *Nucl. Instrum.* **3**, 223–228.
LADELL, J., ZAGOFSKY, A. & PEARLMAN, S. (1975). *J. Appl. Cryst.* **8**, 499–506.
LANGFORD, J. I., BOULTIF, A., AUFFRÉDIC, J. P. & LOUËR, D. (1993). *J. Appl. Cryst.* **26**, 22–33.
LOUËR, D. & LANGFORD, J. I. (1988). *J. Appl. Cryst.* **21**, 430–437.
RIETVELD, H. M. (1969). *J. Appl. Cryst.* **2**, 65–71.
ROSSMANITH, E. (1992). *Acta Cryst.* **48**, 596–610.
ROSSMANITH, E. (1993*a*). *Acta Cryst.* A**49**, 80–91.
ROSSMANITH, E. (1993*b*). *J. Appl. Cryst.* **26**, 753–755.
ROSSMANITH, E., ADIWIDJAJA, G., ECK, J., KUMPAT, G. & ULRICH, G. (1994). *J. Appl. Cryst.* Submitted.
ROSSMANITH, E., WERNER, M., KUMPAT, G., ULRICH, G. & EICHHORN, K. (1993). *J. Appl. Cryst.* **26**, 756–762.
SABINE, T. M. (1987). *J. Appl. Cryst.* **20**, 23–27, 173–178.
WROBLEWSKI, T. (1991). *Acta Cryst.* A**47**, 571–577.

*Acta Cryst.* (1994). A**50**, 68–72

# A Generalized Patterson-Search Method

BY CHRISTER E. NORDMAN

*Department of Chemistry, University of Michigan, Ann Arbor, MI* 48109, *USA*

(*Received* 22 *January* 1992; *accepted* 30 *June* 1993)

## Abstract

A procedure for Patterson search, or molecular replacement, is described in which the criteria of fit are based on matching the asymmetric unit of the entire Patterson function. In rotation search, the Patterson function is compared with the self-Patterson of the search model; in translation search, the comparison is with the full Patterson function of the search model. Significant features of the method are: (1) all overlaps of vector sets of neighboring molecules are taken into account; (2) all overlaps of the search model with neighboring copies are detected and the evaluation bypassed; and (3) the

criteria of fit are flexible and can be expressed in either Patterson or reciprocal space.

## Introduction

In this paper, we consider the task of positioning a known or assumed molecule or molecular fragment in the unit cell of an unknown structure. In general, this task consists of two parts, a rotational part and a translational part. Each is accomplished by systematic examination of the Patterson function of the unknown structure. Many procedures have been devised for carrying out this examination in reciprocal space (Rossmann & Blow, 1962) or in Patterson space (Huber, 1965, 1985; Nordman & Nakatsu, 1963; Braun, Hornstra & Leenhouts, 1969) or in the space of the spherical-harmonic expansion of the Patterson function (Crowther, 1972).

Common to all rotation-search procedures is the neglect of the *predictable* overlap of neighboring vector sets; by this we mean overlap by model vector sets centered on neighboring origins of the Patterson function. Also neglected is the possible interference between the *atoms* of two copies of the model related to one another by a lattice translation.

In this paper, a procedure is described in which the variable and predictable overlap of vector sets is taken into account. A check for interference between adjacent search models is also made and, if detected, the evaluation of that step in the search is bypassed.

The basic premise is that the rotational fit is judged by a comparison of the complete model *self-Patterson* with the Patterson function of the unknown. Translational fit is judged by a similar comparison of the complete model Patterson with the Patterson function of the unknown.

In searches for two or more fragments, the procedure is easily modified: the model self-Patterson (or the complete model Patterson) of fragment (1) can be subtracted from the Patterson function prior to searching for fragment (2).

## The Patterson function

The Patterson function of the unknown can be triclinic, monoclinic or orthorhombic in the present version of the program. It must conform to the symmetry of the space groups $P\bar{1}$, $P2/m$, $C2/m$, $Pmmm$ or $Cmmm$. The stored units of triclinic, monoclinic and orthorhombic Patterson functions are

$$\text{triclinic: } x = 0\text{–}1,\ y = 0\text{–}\tfrac{1}{2},\ z = 0\text{–}1$$

$$\text{monoclinic: } x = 0\text{–}\tfrac{1}{2},\ y = 0\text{–}\tfrac{1}{2},\ z = 0\text{–}1$$

$$\text{orthorhombic: } x = 0\text{–}\tfrac{1}{2},\ y = 0\text{–}\tfrac{1}{2},\ z = 0\text{–}\tfrac{1}{2}.$$

This is consistent with the Fourier-synthesis routines of Ten Eyck (1973), which are used as subroutines. We refer to this as the *natural* Patterson, distinct from the *model* Patterson, as discussed below.

It is assumed that the scale factor and overall thermal parameter of the input data for the natural Patterson are available; these are expected as input to the program. Alternatively, the scale factor and overall anisotropic displacement parameters can be used. An additional sharpening parameter can also be employed.

## Rotation search

The model is placed in the natural cell – not in a large artificial cell – at an arbitrary translational position. Its orientation is arbitrary, unless it possesses rotational symmetry. The model is stepped through the Euler-angle ranges appropriate to the crystal system. If the model possesses $n$-fold rotational symmetry, this symmetry axis is preferably oriented parallel to the axis of the last Euler rotation, reducing the necessary search range by a factor of $n$.

At each step in the Euler-angle search, a check for intermolecular overlap is carried out. This overlap check is concerned with molecules related by *cell translations* only, specifically **a**, **b**, **c**, $\mathbf{a} \pm \mathbf{b}$, $\mathbf{a} \pm \mathbf{c}$, $\mathbf{b} \pm \mathbf{c}$, $\mathbf{a} \pm \mathbf{b} \pm \mathbf{c}$. In centered cells, the appropriate fractional cell translations are used. Overlap is defined here as an intermolecular atom–atom distance less than a specified input parameter, OVLPD. A rapid algorithm is employed, with the execution time essentially proportional to the number of atoms in the model rather than its square. If overlap is detected, the current Euler-angle setting is bypassed and the computation proceeds to the next setting. An index that codes for the offending lattice translation is saved for subsequent display in the output map. The computing time required for the overlap check is negligible compared with the evaluation of the setting.

If no intermolecular overlap is detected, a $P1$ structure-factor calculation is carried out for the model. The resolution is a specifiable input parameter but would normally be the same as the resolution of the natural Patterson. The thermal parameters are isotropic; a provision exists to allow the thermal $B$'s to increase with increasing distance from the center of the model. Tables are used for trigonometric and exponential functions, as well as for scattering factors.

From the $P1$ structure factors, the coefficients for a $P\bar{1}$ Patterson function are obtained, with origin-peak removal optional. The $P\bar{1}$ Patterson function is now computed in the *natural cell*; its asymmetric unit is $x = 0\text{–}1$, $y = 0\text{–}\frac{1}{2}$, $z = 0\text{–}1$. The fractional cell grid is

the same as the fractional cell grid of the natural Patterson. If the natural Patterson is triclinic, the asymmetric unit of the model Patterson is the same as that of the natural Patterson. If the latter is monoclinic, the model asymmetric unit is twice the size of the natural Patterson asymmetric unit and four times if orthorhombic.

The objective is to calculate the true and complete *self-Patterson* of the model, *i.e.* the partial-Patterson function which contains all the vectors within the model and between lattice-translation related copies of the model, but which excludes all vectors dependent on the translational positioning of the model in the cell.

In the triclinic case, this has already been achieved. In the monoclinic case, the model Patterson, $P_M$, is obtained by combining two parts of the $P\bar{1}$ Patterson function into the monoclinic asymmetric unit

$$P_M(x,y,z) = P_{P\bar{1}}(x,y,z) + P_{P\bar{1}}(1-x,y,1-z),$$

*where* $x = 0$–$\frac{1}{2}$, $y = 0$–$\frac{1}{2}$, $z = 0$–1. If the natural Patterson is orthorhombic, the model Patterson is a combination of four parts of the $P\bar{1}$ Patterson function

$$P_M(x,y,z) = P_{P\bar{1}}(x,y,z) + P_{P\bar{1}}(1-x,y,1-z) + P_{P\bar{1}}(1-x,y,z) + P_{P\bar{1}}(x,y,1-z),$$

where $x = 0$–$\frac{1}{2}$, $y = 0$–$\frac{1}{2}$, $z = 0$–$\frac{1}{2}$.

The model self-Patterson is now expressed on the asymmetric unit of the natural Patterson. At the true orientation of the model, the model self-Patterson must be contained in the natural Patterson.

Several criteria of fit are employed and others can easily be added. Some of the criteria of fit are

Sum of products: $\text{SPROD} = \sum(PP_M)$

Correlation coefficient:

$$\text{CORRL} = (N\sum PP_M - \sum P\sum P_M) \times \{[N\sum P^2 - (\sum P)^2]^{1/2}[N\sum P_M{}^2 - (\sum P_M)^2]^{1/2}\}^{-1}.$$

Here, $N$ is the number of grid points in the asymmetric unit or, optionally, all grid points falling within the asymmetric unit and the input parameters RMIN and RMAX, the minimum and maximum Patterson cutoff radii.

Better discrimination is sometimes achieved if a subset of low-$P$ and high-$P_M$ values is emphasized; it will be noted that many of the $P_M$ subunit values will be near zero unless the search model is very large. To accomplish this, the $P$ and $P_M$ values are sorted on $\text{RAT} = P/\text{MAX}(P_M,\langle P_M\rangle)$ and only those $P$ and $P_M$ values that correspond to the lowest values of RAT are considered. The fraction included is governed by input parameters, FRAC. To guard against spuriously negative values of $P_M$ being selected, $P_M$ is replaced by its mean value whenever $P_M < \langle P_M\rangle$. A particularly fast algorithm is used to sort the arrays RAT, $P$ and $P_M$; the sorting time is insignificant. The criteria of fit based on sorting $P$ and $P_M$ are easily modified. Some that have been found useful are

$$\text{W1MAV (FRAC)} = \sum'[(P - \langle P\rangle)P_M]/N$$

$$\text{W2MAV (FRAC)} = \sum'[(P - \langle P\rangle)P_M]/\sum' P_M$$

and

$$\text{W3MAV (FRAC)} = \sum'[(P - \langle P\rangle)P_M^2]/\sum' P_M^2.$$

Here, the sums $\sum'$ contain only those $P$ and $P_M$ values that correspond to the FRAC $\times$ $N$ lowest values of RAT.

The various criteria of fit are printed as functions of the Euler angles. In addition, each is displayed as a 'sigma map', a number field giving $100 \times (C - \langle C\rangle)/\sigma(C)$, where $C$ is the criterion of fit in question. Values less than zero are set to zero. Negative integers indicate a setting rejected for intermolecular overlap; the integer points to the offending lattice translation.

### Translation search

The oriented model is placed in the natural cell and stepwise translated through specified ranges of $x$, $y$ and $z$. As the angular setting presumably has cleared the overlap screening carried out in the rotation search, no checking for overlap of molecules related by lattice translation need be done.

In non-triclinic cells, the checking for overlap between molecules related by space-group-symmetry elements is performed as follows. A set of symmetry matrices is specified in the input, each corresponding to a symmetry element near a possible translational position of the model, whether inside or outside the cell. At present only rotation axes and screw axes are accommodated.

A rapid algorithm is used to detect overlap, as previously described. If overlap is found, *i.e.* an interatomic distance less than the input parameter OLVPD, the translational setting is bypassed and a negative integer flag is saved for display in the sigma map. This negative integer identifies the offending symmetry element.

If no overlap is detected, the structure factors are computed to the desired resolution, normally the same as that of the natural Patterson. An unfactored structure-factor algorithm is used, *i.e.* the structure-factor sum contains all copies of the model present in the cell. Patterson coefficients are computed, with optional origin-peak removal, and the asymmetric unit of the Patterson function calculated. As in the rotation search, several criteria of fit are provided

and the selection can easily be extended. In addition to the functions mentioned earlier, the $R$ factors at low, medium and high resolution, RFCLO, RFCMI and RFCHI, are provided, as well as RFCAL for the entire resolution range. In the sigma maps for $R$ factors, $100 \times (\langle R \rangle - R)/\sigma(R)$ is displayed; values less than zero are set to zero.

## Results and discussion

The procedure has been tested on the Patterson function for metmyoglobin, which is monoclinic with $a = 64.56$, $b = 30.97$, $c = 34.86$ Å, $\beta = 105.86°$, space group $P2_1$. Data to 2.0 Å resolution were used (Takano, 1977). The Patterson function can be computed with an optional sharpening factor, exp(SHRP*STH/L**2); in these tests, SHRP = 0.0 and 20.0 Å$^2$ were used. The origin peak was removed.

Test searches were performed with the program described here (*GENPAT*) and the programs *LATSUM* (Lattman & Love, 1972) and *TRNSUM* (Crowther & Blow, 1967) of the *MERLOT* program package (Fitzgerald, 1988).

The results of these searches are given in Table 1. In each case, the search fragment was taken as the main chain of myoglobin, $C^\beta$'s omitted, a total of 612 atoms. In the rotation search with *GENPAT*, the Patterson radial cutoffs were RMIN = 2.5 and RMAX = 40.0 Å, respectively, but the results were not sensitively dependent on these choices. As far as possible, all adjustable parameters were chosen to be equal to give a valid comparison. Minimum and maximum data resolution were 8.0 and 2.0 Å in all cases. The two-dimensional translation search was performed on a 20 × 10 point uniform grid in $x$ and $z$, with the grid points in each search having a one-to-one correspondence. The rotation searches were performed over a 96-point Eulerian-angle grid, uniformly distributed in the relevant angle space. On account of differences in Cartesian parameter and Eulerian-angle conventions, it was not possible to maintain a one-to-one correspondence of the grid points in the rotation search. In order to allow the *GENPAT* searches to span the entire search range, the overlap distance parameter OVLPD was set to zero. For comparison, the searches were repeated with OVLPD = 2.5 Å, rejecting any setting with an intermolecular interatomic distance less than 2.5 Å. Selected criteria of fit are shown; there were few differences, except that W1MAV, W2MAV and W3MAV depend on the value of FRAC, the fraction of grid points included in the sums.

The false-peak statistics presented in Table 1 suggest that the *GENPAT* searches achieve a somewhat better discrimination than the *MERLOT* searches. *GENPAT* searches of a moderately sharpened (SHRP = 20.0 Å$^2$) Patterson function further improve the results. The reflection data in the *MERLOT* searches were not sharpened and thus, the comparison should be made with the SHRP = 0.0 Å$^2$ *GENPAT* searches. On a setting-by-setting basis, *GENPAT*'s *ROT* is 4.7 times faster than *MERLOT*'s *LATSUM*, whereas *MERLOT*'s *TRNSUM* is 4.1 times faster than *GENPAT*'s *TRN*. Choosing even a conservative value of 2.5 Å for the overlap distance realizes a big advantage in computing time. Clearly, this advantage depends on the dimensions of the search model.

The execution time for the *GENPAT* program is proportional to the number of Euler-angle settings *evaluated*, *i.e.* not bypassed, and is approximately the same for *ROT* and *TRN*. It is also proportional to the number of structure factors calculated at each setting. Finally, it is proportional to the number of atoms in the model fragment, provided it is large. If it is small, a penalty will be paid for inefficient vectorization.

The execution time is essentially independent of the number of Euler-angle settings bypassed for molecular overlap. It is also essentially independent of the density of the Patterson grid and of the number and types of criteria of fit selected. The times

Table 1. *Comparison of results of Patterson searches*

Rotation and translation searches were performed with a 612-atom myoglobin polyglycine main-chain model. The criteria of fit are defined in the text. Peak heights are the number of standard deviations above the mean of the search function. For RFCAL the entry is the number of standard deviations below the mean.

(*a*) Rotation searches

| Program | *GENPAT, ROT* | | *MERLOT, LATSUM* | |
|---|---|---|---|---|
| Criterion of fit | CORRL | CORRL | W3MAV (0.8) | |
| SHRP (Å$^2$) | 20.0 | 0.0 | 0.0 | |
| True orientation | 9.15 | 8.94 | 8.85 | 8.02 |
| Highest false | 0.71 | 0.89 | 0.84 | 1.36 |
| False > 1.00$\sigma$ | 0 | 0 | 0 | 1 |
| False > 0.75$\sigma$ | 0 | 1 | 1 | 7 |
| False > 0.50$\sigma$ | 2 | 5 | 5 | 12 |
| Points evaluated, OVLPD = 0.0 Å | | 96 | | 96 |
| CPU time (s) | | 819 | | 3834 |
| Points evaluated OVLPD = 2.5 Å | | 23 | | |
| CPU time (s) | | 207 | | |

(*b*) Translation searches

| *Program* | *GENPAT, TRN* | | *MERLOT, TRNSUM* | |
|---|---|---|---|---|
| Criterion of fit | CORRL | CORRL | RFCAL | |
| SHRP (Å$^2$) | 20.0 | 0.0 | 0.0 | |
| True position | 11.82 | 11.15 | 12.55 | 9.19 |
| Highest false | 1.19 | 1.45 | 1.32 | 2.36 |
| False > 1.00$\sigma$ | 4 | 9 | 4 | 15 |
| False > 0.75$\sigma$ | 13 | 20 | 7 | 30 |
| False > 0.50$\sigma$ | 28 | 36 | 27 | 51 |
| Points evaluated, OVLPD = 0.0 Å | | 200 | | 200 |
| CPU time (s) | | 1579 | | 384 |
| Points evaluated, OVLPD = 2.5 Å | | 9 | | |
| CPU time (s) | | 82 | | |

given in Table 1 are for an IBM 9021-720 mainframe computer.

The program is in principle available for distribution, but interested parties should consult with the author first.

### References


BRAUN, P. B., HORNSTRA, J. & LEENHOUTS, J. I. (1969). *Philips Res. Rep.* **24**, 85–118.

CROWTHER, R. A. (1972). *The Molecular Replacement Method*, edited by M. G. ROSSMANN, pp. 173–178. New York: Gordon and Breach.

CROWTHER, R. A. & BLOW, D. M. (1967). *Acta Cryst.* **23**, 544–548.

FITZGERALD, P. M. D. (1988). *J. Appl. Cryst.* **21**, 273–278.

HUBER, R. (1965). *Acta Cryst.* **19**, 353–356.

HUBER, R. (1985). *Molecular Replacement*. Proc. of the Daresbury Study Weekend, Daresbury, February 1985, pp. 58–61. SERC Daresbury Laboratory, Warrington, England.

LATTMAN, E. E. & LOVE, W. E. (1972). *Acta Cryst.* **B26**, 1854–1857.

NORDMAN, C. E. & NAKATSU, K. (1963). *J. Am Chem. Soc.* **85**, 353.

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

TAKANO, T. (1977). *J. Mol. Biol.* **110**, 537–568.

TEN EYCK, L. F. (1973). *Acta Cryst.* **A29**, 183–191.


*Acta Cryst.* (1994). A**50**, 72–85

# Space Groups of Trigonal and Hexagonal Quasiperiodic Crystals of Rank 4


BY RON LIFSHITZ AND N. DAVID MERMIN

*Laboratory of Atomic and Solid State Physics, Cornell University, Ithaca, NY* 14853-2501, *USA*


(*Received* 13 *January* 1993; *accepted* 21 *June* 1993)


## Abstract

As a pedagogical illustration of the Fourier-space approach to the crystallography of quasiperiodic crystals, a simple derivation is given of the space-group classification scheme for hexagonal and trigonal quasiperiodic crystals of rank 4. The categories, which can be directly inferred from the Fourier-space forms of the hexagonal and trigonal space groups for periodic crystals, describe general hexagonal or trigonal quasiperiodic crystals of rank 4, which include but are not limited to modulated crystals and intergrowth compounds. When these general categories are applied to the special case of modulated crystals, it is useful to present them in ways that emphasize each of the subsets of Bragg peaks that can serve as distinct lattices of main reflections. These different settings of the general rank-4 space groups correspond precisely to the superspace-group description of (3+1) modulated crystals given by de Wolff, Jannsen & Janner [*Acta Cryst.* (1981), A**37**, 625–636]. As a demonstration of the power of the Fourier-space approach, the space groups for hexagonal and trigonal quasiperiodic crystals of arbitrary finite rank are derived in a companion paper [Lifshitz & Mermin (1994). *Acta Cryst.* A**50**, 85–97].


## I. Introduction

Crystals used to be defined as materials periodic on the atomic scale. As such, they were classified by their *space groups* – subgroups of the full Euclidean group that bring a periodic density into coincidence with itself. Because of the growing numbers and varieties of quasiperiodic crystals, crystals have been redefined* as materials whose diffraction patterns contain Bragg peaks, thereby shifting the essential attribute of crystallinity from position space to Fourier space. A corresponding shift in the crystallographic classification scheme, proposed thirty years ago by Bienenstock & Ewald (1962), has not, however, been widely accepted, probably because they advocated Fourier-space crystallography before quasiperiodic crystals had become of major interest, when there was no strong incentive to make the shift. Now there is.

The conventional extension of the classification scheme to quasiperiodic materials, developed and used by de Wolff, Janssen & Janner (1981) (henceforth JJdW)† to find the 'superspace groups' of (3+1) incommensurately modulated crystals, retains the old criterion of periodicity as the starting point for a crystallographic classification scheme and must therefore treat quasiperiodic structures as three-dimensional sections of structures periodic in a higher-dimensional superspace. The need for such a maneuver is avoided by the Fourier-space classifi-

* Statement of 'terms of reference' of the *ad interim* Commission on Aperiodic Crystals of the International Union of Crystallography.

† See also Yamamoto, Janssen, Janner & de Wolff (1985) and Janssen, Janner, Looijenga-Vos & de Wolff (1992).