*Acta Cryst.* (1964). **17**, 1322

# Patterson Function Interpretation for Molecules Containing Planar Groups

BY P. TOLLIN* AND W. COCHRAN

*Crystallographic Laboratory, Cavendish Laboratory, Cambridge, England*

(*Received* 25 *October* 1963)

A procedure for interpreting the Patterson function of a molecular crystal which contains planar groups of atoms is described, and has been programmed for EDSAC II. Examples of its practical use for structure determination are mentioned. The possibility of more general application of a similar 'sum function' method is discussed, and such methods are compared with a 'minimum function' method originated by Nordman & Nakatsu.

## 1. Introduction

At the beginning of the determination of the structure of a molecular crystal a good deal is often known about the arrangement in space of constituent groups of atoms. This information can obviously be useful in interpreting the Patterson function. Nordman & Nakatsu (1963) have recently described a procedure in which three programs for an IBM 7090 computer are employed in sequence to determine the angular orientation of a rigid group of known internal geometry, the location of this group (in its known orientation) relative to symmetry-related groups in the same unit cell, and finally the determination of the positions of the remaining atoms of the molecule. Their method makes extensive use of the minimum function (Buerger, 1959). We have independently devised a Patterson-interpretation program which is less general in that it can be used only when a considerable proportion of the atoms of the molecule lie in a plane. A preliminary account of this method has already been given at the Munich Symposium (Tollin, 1962). This method is inherently simpler since only a two-dimensional function has to be evaluated in the first stage of the procedure, and because it is based on a sum function rather than a minimum function it does not require storage in the computer of the complete Patterson function at any stage, but can be evaluated as a Fourier series. The method has been programmed for EDSAC II. In § 3 we discuss the possibility of its extension to non-planar groups.

## 2. Description of the method

The orientation of the planar group is described by $\theta, \varphi$, the spherical polar angles which define the normal to the group with respect to the crystal axes, and $\psi$, the azimuthal angle which defines the orientation of a line in the group with respect to a reference line in the plane.

Obviously the vectors between atoms of the group lie in a parallel plane through the origin of the Patterson function. If a disc whose radius is not much greater than the maximum dimension of the molecule is placed with its centre at the origin, then the integral of the Patterson function over the surface of the disc will have a large, and in favourable circumstances a maximum, value when $\theta, \varphi$ of the disc coincide with $\theta, \varphi$ of the planar group. Writing the Patterson function as

$$P(\mathbf{r}) = \sum_{\mathbf{h}} |F(\mathbf{h})|^2 \cos 2\pi \mathbf{h}.\mathbf{r}$$

and defining a function

$$\begin{aligned} t(\mathbf{r}) &= 1 \text{ on the disc} \\ &= 0 \text{ elsewhere,} \end{aligned}$$

we expect

$$\mathscr{I}(\theta, \varphi) = \int_v P(\mathbf{r}) t(\mathbf{r}) dv$$

to be large when the disc is correctly oriented. Apart from constant factors, $\mathscr{I}(\theta, \varphi)$ is identical with

$$I(\theta, \varphi) = \sum_{\mathbf{h}} |F(\mathbf{h})|^2 T(\mathbf{h})$$

where $T$ is the Fourier transform of $t$, and $|F|^2$ is the Fourier transform of $P$. From the definition of $t$ it follows that

$$I(\theta, \varphi) = \sum_{\mathbf{h}} |F(\mathbf{h})|^2 2\pi R^2 \frac{J_1(2\pi R S)}{2\pi R S} \tag{1}$$

where $R$ is the radius of the disc and $S$ is the perpendicular distance of the reciprocal lattice point $\mathbf{h}$ from the normal to the disc.

$I(\theta, \varphi)$ has a value at every point on the surface of a sphere and is conveniently represented on a map which is a projection of this surface. In practice the most convenient projection was found to be the Sanson–Flamsteed sinusoidal equal area projection (Steers, 1949) which was chosen for its ease of construction, its equal-area property, and because it allows a linear interpolation procedure to be used to locate the position of a peak between sampling points, with reasonable accuracy.

---

* Present address: Carnegie Laboratory of Physics, Queen's College, Dundee, Scotland.

An EDSAC II program for the evaluation of $I(\theta\varphi)$ from equation (1) has been written and is described in some detail elsewhere (Tollin, 1963). Several tricks were used to reduce computing time, such as looking up $J_1(x)/x$ in a table constructed so that the value of $S^2$ is used to determine the entry in the table used. This avoids calculation of $J_1(x)$ and extracting a square root, since it is most convenient to calculate $S^2$ from **h** and $\theta$ and $\varphi$. Four separate values of $|F(\mathbf{h})|^2$ were packed in each register, and $S^2$ was expressed as a polynomial in the most frequently varying of $h$, $k$ and $l$, so that as the program moves from one reciprocal lattice point to the next, the minimum amount of computation is required to find the new value of $S^2$. The angle $\theta$ was changed in steps of 2°, and taking

$$\text{Interval in } \varphi = 180/\theta$$

gave a fairly uniform density of sampling points. For crystals of monoclinic symmetry the range to be covered is

$$0 \leq \theta \leq 90°$$
$$0 \leq \varphi \leq 180°$$

and this is reduced for crystals of higher symmetry.

Tests of the program were made with the data for naphthalene (Abrahams, Robertson & White, 1949). It was found to be advantageous to 'sharpen' the data so that the average value of $|F|^2$ was approximately constant in reciprocal space, and to take $R=5$ Å, approximately the maximum dimension of the molecule. Fig. 1(*a*) shows the map obtained with unsharpened $|F|^2$ values and $R=5$ Å; the largest peak had its position in error by not more than 5° in each coordinate. However, in Fig. 1(*b*) is shown the map obtained by using sharpened $|F|^2$ values and $R=5$ Å for the important range of $\theta$; in this case the peak maximum coincides with the position of the plane normal to within less than 1°. From sharpened data for the unknown structure of deoxyadenosine the map shown in Fig. 2 was obtained with $R=2{\cdot}5$ Å. The computing time was about 3 hours. The largest peak which corresponds to the orientation of the purine ring (as was subsequently verified) was in the same position for $R=5$ Å.

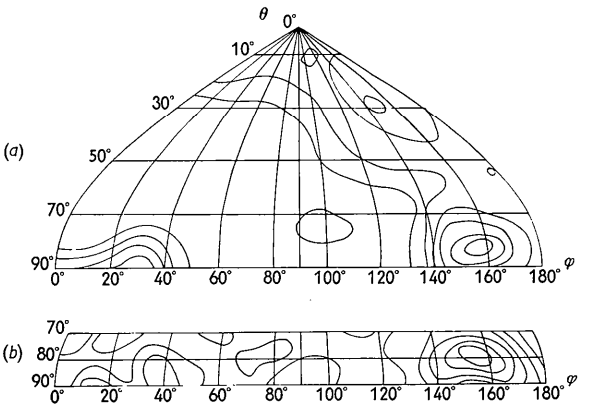

Fig. 1. Naphthalene. (*a*) $I(\theta\varphi)$ $R=5$ Å unsharpened data. (*b*) $I(\theta\varphi)$ for $\theta>70°$, $R=5$ Å, sharpened data.

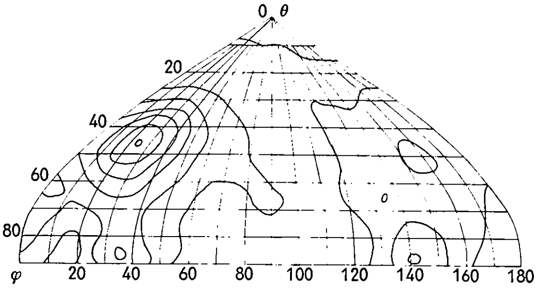

Fig. 2. Deoxyadenosine. $I(\theta\varphi)$ calculated with sharpened data and $R=2{\cdot}5$ Å.

The evaluation and inspection of $I(\theta,\varphi)$ is the first stage of the procedure. The second stage is the evaluation of a plane section of the Patterson function through the origin in the orientation defined by the largest peak of $I(\theta,\varphi)$. (If this section could not be interpreted, sections corresponding to other peaks would be evaluated.) Interpretation of the section consists in superposing on it the vector set of the planar group of atoms, and rotating this until a good fit is obtained. (If the molecular dimensions were accurately known this step could be programmed so that the plane section was not printed out; this could lead to difficulties if the correct peak of $I(\theta,\varphi)$ was not chosen initially.) This determines the angle $\psi$.

In practice two unknown structures have so far been determined by this method, ninhydrin ($C_9H_6O_4$) in which 11 atoms (other than hydrogen) lie in a plane (Tollin, 1963), and the more difficult structure of deoxyadenosine ($C_{10}H_{13}N_5O_3.H_2O$) in which 11 atoms lie in a plane (Sutor, Tollin & Watson, 1964). In both instances the space group is $P2_1$ and the procedure adopted in the final stages was as follows. Let the positions of the $n$ atoms of the planar group, in the correct orientation, with one of them taken as origin, be $\mathbf{r}_j=(x_jy_jz_j)$. The $\mathbf{r}_j$ are therefore all known quantities. The sum function

$$S_n(\mathbf{r}) = \sum_{\mathbf{h}}' |F(\mathbf{h})|^2 \{\sum_j' \cos 2\pi\mathbf{h}.\mathbf{r}_j \cos 2\pi\mathbf{h}.\mathbf{r} + \sum_j \sin 2\pi\mathbf{h}.\mathbf{r}_j \sin 2\pi\mathbf{h}.\mathbf{r}\}$$

was evaluated and printed out. It was then inspected for a group of $n$ peaks having coordinates $(x_j+2x_0, y_j+\frac{1}{2}, z_j+2z_0)$ with identical values of $(x_0, z_0)$, which determines the position of the $2_1$ axis. The determination of the coordinates of the remaining atoms was then made by the usual process of Fourier refinement. Further details are given in the accounts of the individual structures (Tollin, 1963; Sutor, Tollin & Watson, 1964).

## 3. Discussion

The procedure described above can be generalized in several respects but it becomes questionable whether in the general case it has any advantages over the method of Nordman & Nakatsu. When the known group of atoms is not planar, a function of *three* variables, $I(\theta, \varphi, \psi)$, must be evaluated as the first step. The function analogous to the one we have used is

$$\sum_{j,j'=1}^{n} P(\mathbf{r}_j - \mathbf{r}_{j'})$$

where $\mathbf{r}_j$, a function of $(\theta, \varphi, \psi)$, is the coordinate of one atom of the group relative to an arbitrary origin. This sum function can be expressed as a three-dimensional Fourier series in the variables $(\theta, \varphi, \psi)$. The expression is however so cumbersome that it is doubtful whether it could be evaluated more rapidly as a Fourier series, which requires rapid access to a table of about $10^3$ numbers (the values of $|F|^2$), than could be done directly from a table of values of the Patterson function, although this would require rapid access to a table of about $10^5$ numbers. For computers with large fast-access stores it seems clear that the latter method is more efficient, in which case it is certainly better to record against each value of $(\theta, \varphi, \psi)$ the *minimum* value of $P(\mathbf{r}_j - \mathbf{r}_{j'})$ rather than the sum. An exception in favour of the sum function might be made if the dimensions of the known group were not accurately known.

A sum function might find more application at the stage where the rigid group is to be located relative to symmetry-related groups in the same unit cell, since the locating function can be expressed as a relatively simple Fourier series. A procedure analogous to that which was used in the structure determinations of ninhydrin and deoxyadenosine is as follows.

Let $\mathbf{r}_j$ be the coordinate of one of the $n$ atoms of the (correctly oriented) group relative to an arbitrary origin whose position in the unit cell is $\mathbf{r}_0$. Let $T_i$ be the $i$th symmetry operation of the space group and $Z$ the number of equivalent positions. For example for space group $P2_1$, $T_1(xyz) = xyz$, $T_2(xyz) = -x, y+\frac{1}{2}, -z$, and $Z=2$. The sum function obtained by setting down the origin of the Patterson function at each of $n$ sites in each of the first $Z-1$ equivalent positions is

$$S_{n(Z-1)}(\mathbf{r}) = \sum_{i=1}^{Z-1} \sum_{j=1}^{n} P(\mathbf{r} - T_i(\mathbf{r}_j + \mathbf{r}_0))\,.$$

This expression is now summed over the remaining $n$ points for which $\mathbf{r} = T_Z(\mathbf{r}_{j'} + \mathbf{r}_0)$ to give a locating function

$$\begin{aligned} Q(\mathbf{r}_0) &= \sum_{j'=1}^{n} \sum_{i=1}^{Z-1} \sum_{j=1}^{n} P(T_Z(\mathbf{r}_{j'} + \mathbf{r}_0) - T_i(\mathbf{r}_j + \mathbf{r}_0)) \\ &= \sum_{\mathbf{h}} \sum_{j'=1}^{n} \sum_{i=1}^{Z-1} \sum_{j=1}^{n} |F(\mathbf{h})|^2 \exp\,[2\pi i \mathbf{h}\,.\{T_i(\mathbf{r}_j + \mathbf{r}_0) \\ &\qquad - T_Z(\mathbf{r}_{j'} + \mathbf{r}_0)\}] \end{aligned} \qquad (2)$$

which is a Fourier series in the variable $\mathbf{r}_0$ and should have a peak at the correct value of $\mathbf{r}_0$. When $\mathbf{r}_0$ has its correct value, the value of $Q(\mathbf{r}_0)$ is simply the sum of the values of the Patterson function at positions which correspond to vectors between atoms in different groups. As an example, when the space group is $P2_1$, equation (2) reduces to

$$\begin{aligned} Q(x_0, z_0) = \sum_{hkl} (-1)^k |F(hkl)|^2 \sum_{j,j'=1}^{n} \cos 2\pi (h(x_j + x_{j'} + 2x_0) \\ + k(y_j - y_{j'}) + l(z_j + z_{j'} + 2z_0))\,. \end{aligned}$$

One of us (P.T.) is indebted to the Steetley Magnesite Company for the award of a Research Bursary, during the tenure of which the work reported here was carried out. We wish to thank Dr M. V. Wilkes for permission to use EDSAC II.

## References

ABRAHAMS, S. C., ROBERTSON, J. M. & WHITE, J. C. (1949). *Acta Cryst.* **2**, 233.

BUERGER, M. J. (1959). *Vector Space*. New York: Wiley.

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

STEERS, J. A. (1949). *The Study of Map Projections*. London: Univ. London Press.

SUTOR, D. J., TOLLIN, P. & WATSON, D. G. (1964). To be published.

TOLLIN, P. (1962). Paper presented at the Symposium: 'Recent Advances in the Experimental and Theoretical Methods of Crystal Structure Research': Munich.

TOLLIN, P. (1963). Ph. D. Thesis, Univ. of Cambridge.