

$$I = \int_a^b \mathcal{J}_0^2 \{ B[(x_q - a)(b - x_q)]^{1/2} \} dx_q. \quad (B.1)$$

Strictly speaking, the parameter B depends on x_q ; however, both the polarization factor C and $|\chi_h|/\lambda^2$ can be assumed constant within the integration limits if condition (9) is satisfied. Assuming $B = \text{constant}$, we calculate the integral in expression (B.1) following Kato's (1961*b*) method. Let $\mathcal{J}_0^2(\zeta)$ be in the form of a series:

$$\mathcal{J}_0^2(\zeta) = \sum_{m=0}^{\infty} (-1)^m \frac{1}{(m!)^2} \frac{(2m-1)!!}{(2m)!!} \zeta^{2m}. \quad (B.2)$$

Using decomposition (B.2) and having made a change of variables in (B.1): $u = (x_q - a)/(b - x_q)$, we have:

$$I = B^{-1} \sum_{m=0}^{\infty} (-1)^m \frac{1}{(m!)^2} \frac{(2m-1)!!}{(2m)!!} [B(b-a)]^{2m+1} \times \int_0^{\infty} \frac{u^m}{(1+u)^{2m+2}} du. \quad (B.3)$$

The integral in (B.3) is the tabulated (Dwight, 1961):

$$\int_0^{\infty} \frac{u^m}{(1+u)^{2m+2}} du = \frac{(m!)^2}{(2m+1)!}. \quad (B.4)$$

Substituting (B.4) into (B.3) we have the expression

which can be written:

$$I = B^{-1} \int_0^{B(b-a)} \mathcal{J}_0(\varrho) d\varrho. \quad (B.5)$$

Here

$$B(b-a) = (KC|\chi_h|t/\cos\theta)(1-t^2/D_3^2)^{-1/2} = 2A(1-t^2/D_3^2)^{-1/2} \quad (B.6)$$

where A is the parameter of the dynamical theory of X-ray diffraction introduced by Zachariasen (1945). Since we assumed condition (9) to be satisfied, the upper limit of integration in (B.5) is equal to $2A$, $B \simeq KC|\chi_h|/2 \sin\theta$, and (B.5) corresponds to Waller's integral.

References

- ARISTOV, V. V., SHMYTKO, I. M. & SHULAKOV, E. V. (1977). *Acta Cryst.* A33, 412–418.
 DWIGHT, H. B. (1961). *Tables of Integrals and Other Mathematical Data*. New York: Macmillan.
 HART, M. & LANG, A. R. (1965). *Acta Cryst.*, 19, 73–77.
 HATTORI, H., KURIYAMA, H. & KATO, N. (1965). *J. Phys. Soc. Japan*, 20, 1047–1050.
 KATO, N. (1961*a*). *Acta Cryst.* 14, 526–532.
 KATO, N. (1961*b*). *Acta Cryst.* 14, 627–636.
 WALLER, I. (1926). *Ann. Phys.* 79, 261–273.
 ZACHARIASEN, W. H. (1945). *Theory of X-ray Diffraction in Crystals*. Sec. III, 8. New York: John Wiley.

Acta Cryst. (1977). A33, 423–425

The Use of Structure Factors to find the Origin of an Oriented Molecular Fragment*

BY A. D. RAE

School of Chemistry, University of New South Wales, Kensington, NSW 2033, Australia

(Received 7 July 1976; accepted 1 November 1976)

When a molecular fragment has a known orientation, it is possible to test rapidly all possible positions in the unit cell for this fragment by calculating structure factors at each position for a moderately large number of reflexions. A program, *LOCOFOROM*, has been written for this purpose and application to two structures is discussed.

Introduction

The location of a known molecular fragment in a crystal often takes place in two distinct stages. The first, and usually the easiest, stage is the determination of the orientation of the fragment. Indeed, it has long been recognized that a consistent but incorrect set of phases from a statistical model often yields an E map

with recognizable molecular fragments of correct orientation but wrong position.

The orientation of a fragment may often be unambiguously obtained from the Patterson function (Nordman, 1970) or from a comparison of a calculated transform for a fragment with the observed reflexions (Tollin & Cochran, 1964). This information is sometimes used to assist statistical phase-determination methods (Thiessen & Busing, 1974).

The second stage is the determination of the positions in the crystal for these fragments. However, difficulties are often encountered at this stage and the

* Research performed at Oak Ridge National Laboratory and sponsored by the Energy Research and Development under contract with Union Carbide Corporation.

point of this paper is to show that it is not impractical to compare observed and calculated values of the structure factor for a systematic testing of possible sites for a fragment on a grid of 0.2 to 0.3 Å sampling frequency. Practical experience with small molecules suggests that this fine a grid is essential.

The concept of the method described in this paper is mentioned but not extensively developed by Stout & Jensen (1968). Apart from certain space-group-specific applications using projection data (Taylor & Morley, 1959; Bhuiya & Stanley, 1964), little effort seems to have been made to develop the method as a general procedure with three-dimensional diffraction data for any space group. This is understandable as few successful solutions could be expected from projection data because of an insufficient number of high-intensity data and the overlap of atoms in projection (some overlapping atoms possibly being omitted from a trial fragment). Our results also show that the use of too coarse a grid (say 0.4 to 0.5 Å) for three-dimensional data would also probably fail to give a solution. Points 0.2 to 0.3 Å away from a correct solution show no better statistics than a great many other points. However, as shown in this paper, the computing cost of doing the job properly is not great.

An inherent advantage in the approach described in this paper compared with the use of a translation function (Karle, 1972) is that there is no restriction on the choice of function of observed and calculated structure factor amplitudes to determine best the position of a fragment. Also, more than one such function can be simultaneously tested.

Theory

If (θ_m, \mathbf{d}_m) is a symmetry operation where $(\theta_m, \mathbf{d}_m)\mathbf{r} = \theta_m\mathbf{r} + \mathbf{d}_m$, then the structure factor may be described as $F(\mathbf{S}) = \sum_m F_m(\mathbf{S})$ where

$$F_m(\mathbf{S}) = \exp(2\pi i \mathbf{d}_m \cdot \mathbf{S}) \sum_n a_n f_n(\mathbf{S}) \psi_n(\theta_m^{-1}\mathbf{S})$$

where a_n is the occupancy factor, $f_n(\mathbf{S})$ is the scattering factor and $\psi_n(\mathbf{S})$ is the Fourier transform of the probability density function for the position of the n th atom in the asymmetric unit. If all atoms in the asymmetric unit are moved by a fixed amount from positions \mathbf{r}_n to positions $\mathbf{r}_n + \mathbf{r}_o$ then $F_m(\mathbf{S})$ is transformed to $\exp(2\pi i \mathbf{r}_o \cdot \theta_m^{-1}\mathbf{S}) F_m(\mathbf{S})$. We thus see that it is not necessary to evaluate $F_m(\mathbf{S})$ for every point \mathbf{r}_o that we wish to test, but rather simply to evaluate $F_m(\mathbf{S})$ for the point $\mathbf{r}_o = 0$.

If \mathbf{r}_o is only allowed to have certain specific values $(n_1/N)\mathbf{a}_1 + (n_2/N)\mathbf{a}_2 + (n_3/N)\mathbf{a}_3$ where the \mathbf{a}_i are axial lengths, N is a fixed integer and n_1, n_2, n_3 are integers which are stepped to sample three-dimensional space, then the allowed values of $\exp(2\pi i \mathbf{r}_o \cdot \theta_m^{-1}\mathbf{S})$ can be obtained by a simple table look up with no need for interpolation. Initial evaluation of $\mathbf{a}_i \cdot \theta_m^{-1}\mathbf{S}$ for a reflexion whose scattering vector is \mathbf{S} allows simple

modulo arithmetic logic to ascertain which elements in pre-evaluated cosine and sine tables to look up for any particular grid point. Points are tested one layer at a time. For each layer, data for one reflexion is read, and F_c evaluated for each grid point before the next reflexion is read.

The maximum volume that needs to be searched is not the volume of one asymmetric unit. Rather it is the volume defined by the choice of origin for the space group (Hirshfeld, 1968), e.g. $\frac{1}{8}$ of the unit cell for $P\bar{1}$, $\frac{1}{4}$ of the unit cell for $P4_12_12_1$, $\frac{1}{4}$ of the section $y=0$ for $P2$.

Discussion

A program, *LOCOFOROM*, to locate the origin for an oriented molecule has been written at Oak Ridge National Laboratory and incorporates the above ideas. The first test of the program was with data for imidazole-4-acrylic acid dihydrate ($C_6H_{10}O_4N_2$) which crystallizes in space group $P2_12_12_1$ [$a=9.624(1)$, $b=12.881(1)$, $c=6.714(1)$ Å at 22°C, $Z=4$]. From an incorrect phase set approximate relative coordinates for the atoms of imidazole-4-acrylic acid could be found from an E map. However, the molecule was in the wrong position and the relative positions of the two water molecules of crystallization could not be found. The 271 reflexions of highest $|F|$ were used and the volume searched was between fractional coordinates $0-\frac{1}{2}$, $0-\frac{1}{2}$, $0-\frac{1}{2}$ in steps of $3/120$, $2/120$, and $4/120$ in x, y, z respectively.

Values of $R_1 = \sum |\Delta F| / \sum F_o$, $R_2 = \sum (\Delta F)^2 / \sum F_o^2$, and $R_3 = \sum |F_o|^2 |F_c|^2 / \sum |F_o^4|$ were plotted and calculation took less than 3 min on an IBM 360/91. The correct position should presumably correspond to minimum values of R_1, R_2 and a maximum value of R_3 . Molecular coordinates were taken from the incorrectly phased E map and no attempt was made to idealize the geometry of the ten-nonhydrogen-atom fragment. The incorrect origin ($\mathbf{r}_o = 0$) gave $R_1 = 0.45$, $R_2 = 0.28$, $R_3 = 1.35$. The scale and a simple overall isotropic temperature parameter were obtained from a Wilson plot and a constrained least-squares refinement program *RAELS* was used to provide the structure-factor information compatible with the program *LOCOFOROM*.

A correct structure-factor refinement (Hawkinson, 1977) was obtained from the point $\mathbf{r}_o = (15/120, 0/120, 24/120)$ which gave values of $R_1 = 0.42$, $R_2 = 0.25$, $R_3 = 1.36$. Values of R_1 and R_2 for this point were the best values for any of the tested points, though the best value of R_3 was 1.43. R_3 proved to be an imprecise function for testing purposes despite the fact that it is commonly used for determining fragment orientation. The second best value of $R_1 + R_2$ was for the point $\mathbf{r}_o = (15/120, 0/120, 0/120)$ with $R_1 = 0.43$, $R_2 = 0.27$, $R_3 = 1.38$. A correct refinement could also be obtained from this point. The point $\mathbf{r}_o = 0$ corresponds to an almost planar molecule lying perpendicular to c at $z = 0.65$. The two best solutions then have the

same x, y coordinates for atoms but with $z \approx 0.85$ or 0.65 . As a consequence, the transformation $x, y, z \rightarrow x, y, z - 0.2$ closely approximates the transformation $x, y, z \rightarrow x, y, 1.5 - z$ ($z \approx 0.85$) which is a transformation that changes the correct structure into its enantiomorph.

The second test of the program was with the 15-nonhydrogen-atom compound norcubebanone $C_{14}H_{20}O$ [$a = 8.480$ (2), $b = 23.502$ (5), $c = 6.314$ (1) Å at 22°C , space group $P2_12_12_1$, $Z = 4$]. An eight-atom fragment (a half-chair cyclohexane ring with two attached C atoms) was used. The geometry was idealized from preconceptions [not the known structure, (Thiessen, 1977)] and the orientation of this fragment was found with the program *ORTRAN* (Thiessen & Busing, 1974). The 177 reflexions with $E > 1.5$ were used. The volume searched was between fractional coordinates $0 - \frac{1}{2}$, $0 - \frac{1}{2}$, $0 - \frac{1}{2}$ in steps of $3/100$, $1/100$, and $4/100$ in x , y , z respectively. The correct position of the fragment again corresponded to the best value of R_1 , namely $R_1 = 0.47$. The value of R_2 was 0.257 compared with a best value of 0.250 and R_3 was 0.432 compared with a best value of 0.440 . The use of R_3 as a criterion of correctness was again poor; for example, the correct position corresponded to the third best value of $R_1 + R_2 - R_3$. The value of R_1 for the position with the best value of $R_1 + R_2 - R_3$ was $R_1 = 0.49$, while for the second best point R_1 was 0.48 . Only these three points stood out as possible solutions. If values of $R_1 + R_2$ were considered, the correct solution was again the best point (value 0.73 as against 0.74 and 0.75). The time for the calculation of the map was just over 2 min on the IBM 360/91.

Obviously, further thought can be given as to what data to select for the calculation and what functions of the structure factor are best able to distinguish a correct solution. However, plotting a number of functions does not add greatly to the computing time. Also, a finer search about promising points can be undertaken with more data.

To locate simultaneously two fragments in an asymmetric unit, a six-dimensional space would have to

be searched. This makes the proposition prohibitively expensive. However, it should be pointed out that packing arguments could be used to greatly reduce the volume that needs to be searched. Of course, if one fragment is known in both orientation and position, a second fragment can be located with the ideas expounded in this paper by simply including a component of the structure factor which is invariant for various sampling points r_o . Two fragments in the triclinic space group $P1$ is of course a case that is always amenable to this approach. The first fragment can always be fixed at the origin of the unit cell.

I wish to acknowledge W. E. Thiessen of Oak Ridge National Laboratory and S. W. Hawkinson of the University of Tennessee for supplying intensity data, and for taking an active interest in this project, even though they had each already solved the above troublesome structures by other means.

References

- BHUIYA, A. K. & STANLEY, E. (1964). *Acta Cryst.* **17**, 746–748.
- HAWKINSON, S. W. (1977). *Acta Cryst.* To be published.
- HIRSHFELD, F. L. (1968). *Acta Cryst.* **A24**, 301–311.
- KARLE, J. (1972). *Acta Cryst.* **B28**, 820–824.
- NORDMAN, C. E. (1970). *Crystallographic Computing*, Topic B, pp. 81–123, containing contributions from R. A. JACOBSEN, N. C. SEEMAN, P. TOLLIN, R. HUBER, J. HORNSTRA, C. E. NORDMAN & J. W. SCHILLING. Edited by F. R. AHMED, S. R. HALL & C. P. HUBER. Copenhagen: Munksgaard.
- STOUT, G. H. & JENSEN, L. H. (1968). *X-ray Structure Determination*, pp. 309–310. London: Macmillan.
- TAYLOR, C. A. & MORLEY, K. A. (1959). *Acta Cryst.* **12**, 101–105.
- THIESSEN, W. E. (1977). *Acta Cryst.* To be published.
- THIESSEN, W. E. & BUSING, W. R. (1974). *Acta Cryst.* **A30**, 814–821.
- TOLLIN, P. & COCHRAN, W. (1964). *Acta Cryst.* **17**, 1322–1324.