

Molecular Location from Minimum Residual Calculation

BY A. K. BHUIYA AND E. STANLEY

Department of Physics, College of Science and Technology, Manchester, 1, England

(Received 25 July 1963)

When the orientation of a molecule is known its position in the unit-cell projection can be determined by calculating the value of the residual as the molecule is moved systematically over the unit cell. The position of the minimum value of the residual gives the best position for the molecule. With fast computers the calculation is quite short and leads to the correct structure.

Introduction

When the stereochemistry of the molecule is well known, or can be reasonably well predicted, as is often the case with organic molecules, it is possible to divide the structure determination into two distinct steps. The first step is to determine the orientation of the molecule with respect to the crystal axes; the second is to locate the position of the molecule in the unit cell. The first step can often be carried out satisfactorily by optical transform methods (Hanson, Lipson & Taylor, 1953) or from a study of the Patterson peaks near the origin (e.g. Cochran, 1951). One method of locating the molecule whose orientation is known has been given by Taylor (1954) and in a different form by Taylor & Morley (1959). The methods make use of a small number of very weak reflexions which lie in strong regions of the Fourier transform of a single molecule. A more ambitious method for determining both the orientation and the location of the molecules in the unit cell has been suggested by Milledge (1962).

The method of molecular location described in this paper makes use of much more of the available data, and is objective in that its success does not depend on the choice of reflexions.

Principles

When the orientation of the molecule has been determined, the best position in the unit cell is where the observed and calculated structure factors are in best agreement. The popular criterion for judging the correctness of a structure, and the one most familiar to structural crystallographers, is the value of $R = \sum ||F_o| - |F_c|| / \sum |F_o|$. There are other possible agreement criteria which have been suggested from time to time (e.g. Booth, 1945) which may have more theoretical justification but appear to have no great advantage in the present work. The value of R is calculated, with the molecule in fixed orientation with respect to the crystal axes, at all positions in the unit-cell projection and is plotted as a function of the coordinates. The position of the minimum value of R should give the position of the molecule in the unit-cell projection. If more than a very few reflexions are used this is a time-consuming process and computer

techniques have been adopted. The calculation can be simplified by splitting it up into two parts, one involving the coordinates of the atoms of the molecule relative to an arbitrary origin, the other involving the position of this origin relative to the unit-cell projection origin.

A brief outline of the procedure for the plane group pgg is given and trivial alterations only would be necessary to enable it to be applied to other plane groups.

If x_i, y_i are the fractional coordinates of the i th atom of the projected molecule with respect to some arbitrary origin, and X, Y are the fractional coordinates of this origin with respect to the origin of the unit-cell projection, the structure factor can be written

$$\left. \begin{aligned} F(hk) &= 4 \sum_i f_i \exp(-B \sin^2 \theta / \lambda^2) \cos 2\pi h(X + x_i) \\ &\quad \times \cos 2\pi k(Y + y_i) \quad \text{when } h+k=2n \\ F(hk) &= 4 \sum_i f_i \exp(-B \sin^2 \theta / \lambda^2) \sin 2\pi h(X + x_i) \\ &\quad \times \sin 2\pi k(Y + y_i) \quad \text{when } h+k=2n+1 \end{aligned} \right\} (1)$$

where f_i is the atomic scattering factor of the i th atom and B is a general isotropic temperature factor. Equation (1) can be rewritten as

$$\begin{aligned} F_{hk}(X, Y) &= 4\{A \cos 2\pi hX \cos 2\pi kY - B \cos 2\pi hX \sin 2\pi kY \\ &\quad - C \sin 2\pi hX \cos 2\pi kY + D \sin 2\pi hX \sin 2\pi kY\} \end{aligned}$$

when $h+k=2n$, and

$$\begin{aligned} F_{hk}(X, Y) &= 4\{A \sin 2\pi hX \sin 2\pi kY + B \sin 2\pi hX \cos 2\pi kY \\ &\quad + C \cos 2\pi hX \sin 2\pi kY + D \cos 2\pi hX \cos 2\pi kY\} \end{aligned}$$

when $h+k=2n+1$, where

$$\begin{aligned} A &= \sum_i f_i \exp(-B \sin^2 \theta / \lambda^2) \cos 2\pi h x_i \cos 2\pi k y_i, \\ B &= \sum_i f_i \exp(-B \sin^2 \theta / \lambda^2) \cos 2\pi h x_i \sin 2\pi k y_i, \\ C &= \sum_i f_i \exp(-B \sin^2 \theta / \lambda^2) \sin 2\pi h x_i \cos 2\pi k y_i, \\ \text{and } D &= \sum_i f_i \exp(-B \sin^2 \theta / \lambda^2) \sin 2\pi h x_i \sin 2\pi k y_i. \end{aligned}$$

A, B, C, D may be regarded as the components of the molecular scattering factor appropriate to this plane group. A, B, C and D depend only on the molecular model and the temperature factor and are constant for any particular model.

The calculation proceeds as follows:

- (1) The values of $f_i \exp(-B \sin^2 \theta/\lambda^2)$ are calculated for each reflexion and each type of atom, the analytical f values used being calculated by use of the constants given by Forsyth & Wells (1959) and the value of B being found by the method of Wilson (1942). These values of

$$f_i \exp(-B \sin^2 \theta/\lambda^2)$$

are stored temporarily.

- (2) The values of A, B, C and D are calculated for each reflexion and are stored permanently, and the values of $|F_o|$ are also calculated and stored permanently.
- (3) The values of $F_{hk}(X, Y)$ are calculated for each reflexion for a particular value of X and Y and the corresponding value of $R(X, Y)$ is calculated and printed. The values of X and Y are varied systematically in increments so chosen to be fine enough to make interpolation of the minimum value reasonable and coarse enough to keep the calculation as short as possible. Using the Mercury computer and only the fast access store it is possible to accommodate a group of 40 reflexions which can be chosen in any way. The time taken to calculate and punch each R value is $2\frac{1}{2}$ seconds and the total time for a 900 point calculation including input time is 40 minutes.

Usually the molecular model which is used will be an inaccurate one. For this reason it is better to use only the low angle reflexions which are relatively insensitive to small errors in coordinates. The strong

reflexions may be of more use than the weak ones because a strong reflexion can only arise if the molecular Fourier transform is strong at that point in reciprocal space and if the molecules are in positions where the contributions reinforce each other. A weak reflexion can arise *either* because the molecular Fourier transform is weak at that point *or* because the molecules are in positions where their contributions cancel. Taylor (1954) and Cochran & Douglas (1957) regard the weak reflexions as a better criterion; there is no doubt that the weak reflexions are more sensitive. $R(X, Y)$ maps have been calculated with the use of (a) strong reflexions only, (b) weak reflexions only and (c) all reflexions regardless of magnitude. Another possible advantage of using the strong reflexions only is that the R values obtained do show the kind of variation which might help to distinguish a reasonable structure from a random one.

Applications

The method has been applied to several problems. Two examples are given, one of a known structure, the other unknown.

(a) Triphenylene

This structure was first investigated by Klug (1950) and was later redetermined by Pinnock, Taylor & Lipson (1956), who found a different position for the molecule but the same orientation. The relative coordinates of Klug's molecule were used to calculate the $R(X, Y)$ map from 40 low-order strong reflexions. The map is shown in Fig. 1. The lowest value of the residual is 0.20 at the position 0.453, 0.081. This is sensibly the same position as that found by Pinnock, Taylor & Lipson (1956), 0.458, 0.079. It is interesting to note that there is a minor minimum ($R=0.34$) at 0.425, 0.131, the position found by Klug.

(b) Naphthocinnoline

This compound, which is at present being studied by the authors, has the space group $Fdd2$ with sixteen molecules in the unit cell. The [001] projection has the plane group pgg . The approximate orientation of the molecule was determined by the optical transform method. The molecular location technique of Taylor & Morley (1959) suggested several possible molecular positions, none of which gave a structure which would

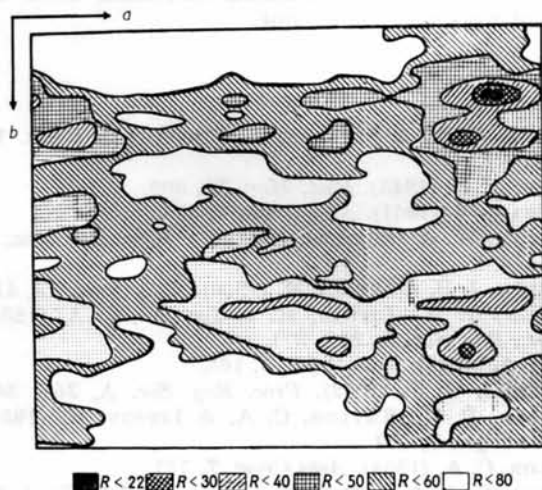


Fig. 1. $R(X, Y)$ map for triphenylene using Klug's model and 40 strong low-order reflexions.

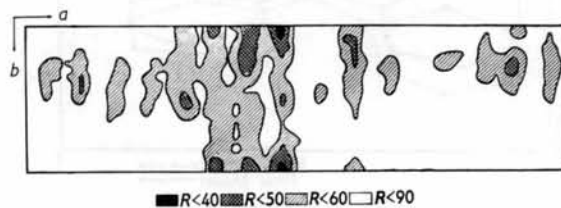


Fig. 2. $R(X, Y)$ map for naphthocinnoline using 40 strong low-order reflexions.

refine to a residual less than 0.35. It is possible that a more careful choice of reflexions would have given better results.

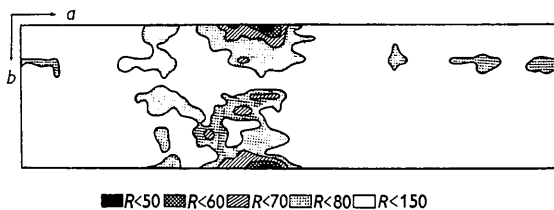


Fig. 3. $R(X, Y)$ map for naphthocinnoline using 40 weak low-order reflexions.

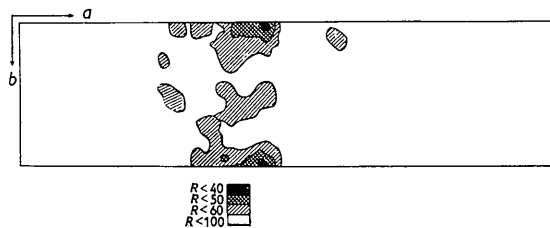


Fig. 4. $R(X, Y)$ map for naphthocinnoline using 40 lowest-order reflexions.

The molecular position was then determined by the method described in this paper. Forty strong, low order, reflexions were used. The $R(X, Y)$ map is shown in Fig. 2. The principal minimum is at 0.233, 0.013 where the residual has the value 0.34. (This position is fairly close to one of the least prominent minima obtained in the method of Taylor & Morley.) The calculation was repeated with 40 low order weak reflexions. The $R(X, Y)$ map is shown in Fig. 3. The position of the minimum at 0.233, 0.000 is sensibly the same as that obtained with the strong reflexions although, of course, the value of the minimum is much larger. The $R(X, Y)$ map of the 40 lowest order reflexions irrespective of magnitude is shown in Fig. 4. The principal minimum is, of course, in the same position but it is rather more extended. Taking this position as the starting point the structure has been successfully refined by the method of Bhuiya & Stanley (1963) to the present R value of 0.16. Some

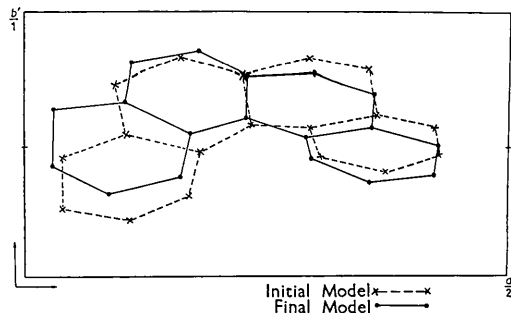


Fig. 5. The initial model of naphthocinnoline in its original orientation is the position given by this method of molecular location compared with the final refined model.

idea of the very approximate nature of the model can be seen from the comparison of the original model used for the calculation of the $R(X, Y)$ maps and the final refined model shown in Fig. 5.

Discussion

During the trial stages of the method it has been observed that both the strong reflexions only and the weak reflexions only give a single prominent minimum at the same position provided the model is substantially correct. The subsidiary minima are often at quite different places on the two maps. If the orientation of the model is quite wrong, experience indicates that there will be no very prominent minimum on either of the two maps and there may be no common minimum; in addition the general level of the R values will be high. These characteristics are quite useful in indicating when a model is not acceptable. If the model is correct the map of $R(X, Y)$ for the low order reflexions irrespective of magnitude will show only one prominent minimum.

It seems that the best procedure is (i) to calculate the $R(X, Y)$ map from the low order strong reflexions only and then, if the map shows a prominent minimum and the general level of the R values is small, (ii) to calculate the $R(X, Y)$ map from only the weak reflexions. One of the minima should coincide with the principal minimum on the first map and this position is taken as the starting point for refinement.

We are grateful to Professor H. Lipson for his interest in this work, to Dr. S. S. Hasan and Mr. P. Main for helpful discussions and to Professor F. C. Williams for providing the facilities of the Manchester University Computing Laboratory. One of us (A.K.B.) is indebted to the Pakistan Council of Scientific and Industrial Research for study leave and to the Government of Pakistan for financial assistance while the present work was carried out.

References

- BHUIYA, A. K. & STANLEY, E. (1963). *Acta Cryst.* **16**, 981.
 BOOTH, A. D. (1945). *Phil. Mag.* **36**, 609.
 COCHRAN, W. (1951). *Acta Cryst.* **4**, 376.
 COCHRAN, W. & DOUGLAS, A. (1957). *Proc. Roy. Soc. A*, **243**, 281.
 FORSYTH, J. B. & WELLS, M. (1959). *Acta Cryst.* **12**, 412.
 HANSON, A. W., LIPSON, H. & TAYLOR, C. A. (1953). *Proc. Roy. Soc. A*, **218**, 371.
 KLUG, A. (1950). *Acta Cryst.* **3**, 165.
 MILLEDGE, H. J. (1962). *Proc. Roy. Soc. A*, **267**, 566.
 PINNOCK, P. R., TAYLOR, C. A. & LIPSON, H. (1956). *Acta Cryst.* **9**, 173.
 TAYLOR, C. A. (1954). *Acta Cryst.* **7**, 757.
 TAYLOR, C. A. & MORLEY, K. A. (1959). *Acta Cryst.* **12**, 101.
 WILSON, A. J. C. (1942). *Nature, Lond.* **150**, 152.