

Although this solution is geometrically impossible (compare Schütte & van der Waerden, 1951) it is a useful approximation in terms of which to describe the observed coordination.

References

- BENNET, M. J., COTTON, F. A., LEGZDINS, P. & LIPPARD, S. J. (1968). *Inorg. Chem.* **7**, 1770.
 BURNS, J. (1969). Private communication.
 BUSING, W. R., MARTIN, K. O. & LEVY, H. A. (1962). *ORFLS*. Report ORNL-TM-305, Oak Ridge National Laboratory, Tennessee.
 CHEN, W. (1967). Dissertation, Louisiana State University.
 COTTON, F. A. & WISE, J. J. (1966). *Inorg. Chem.* **5**, 1200.
 COTTON, F. A. & WOOD, J. S. (1964). *Inorg. Chem.* **3**, 245.
 DUNITZ, J. D. & ORGEL, L. E. (1960). *Advanc. Inorg. Chem. Radiochem.* **2**, 1.
 DUNNING, A. J. & VAND, V. (1969). *Acta Cryst.* **A25**, 489.
 EISENTRAUT, K. J. & SIEVERS, R. E. (1965). *J. Amer. Chem. Soc.* **87**, 5254.
 ERASMUS, C. S. & BOEYENS, J. C. A. (1969). *Acta Cryst.* **A25**, S162.
 GILLESPIE, R. J. (1960). *Canad. J. Chem.* **38**, 818.
 HUSH, N. S. (1962). *Aust. J. Chem.* **15**, 378.
International Tables for X-ray Crystallography (1965). 2nd ed. Vol. I, p. 99. Birmingham: Kynoch Press.
 LINGAFELTER, E. C. (1966). *Coord. Chem. Rev.* **1**, 151.
 LINGAFELTER, E. C. & BRAUN, R. L. (1966). *J. Amer. Chem. Soc.* **88**, 2951.
 MODE, V. A. & SMITH, G. S. (1969). *J. Inorg. Nucl. Chem.* **31**, 1857.
 SCHÜTTE, K. & VAN DER WAERDEN, B. L. (1951). *Math. Annalen*, **123**, 96.
 SCHWARBERG, J. E., GERE, D. R., SIEVERS, R. E. & EISENTRAUT, K. J. (1967). *Inorg. Chem.* **6**, 1933.

Acta Cryst. (1970). **B26**, 1854

A Rotational Search Procedure for Detecting A Known Molecule In a Crystal*

BY EATON E. LATTMAN AND WARNER E. LOVE

Thomas C. Jenkins Department of Biophysics, Johns Hopkins University, Charles and 34th Streets, Baltimore, Maryland 21218, U.S.A.

(Received 20 August 1969)

A computationally swift modification of the Rossmann-Blow rotation function has been developed. With the use of this function the single chains present in the D2 crystals of hemoglobin from the sea lamprey, *Petromyzon marinus*, have been shown to resemble the sperm-whale metmyoglobin molecule, and the orientation of the lamprey hemoglobin molecules in the unit cell of this crystal has been found. The results are confirmed by the crystal structure analysis of lamprey hemoglobin.

Introduction

The oxygen-carrying heme proteins from a variety of sources have similar tertiary structures, but occur in widely differing crystal forms. In particular, the α - and β -chains of horse and human hemoglobin (Cullis, Muirhead, Perutz, Rossmann & North, 1962; Muirhead, Cox, Mazzarella & Perutz, 1967), the single chains of the hemoglobins from the common blood-worm, *Glycera dibranchiata* (Padlan & Love, 1968) and from the larval form of the fly, *Chironomus thummi*, (Huber, Formanek & Epp, 1968), and the single chains of seal and sperm-whale myoglobin (Scouloudi, 1969; Bodo, Dintzis, Kendrew & Wyckoff, 1959) all appear to have essentially the same topology when viewed at low resolution, although the crystallographic arrangements in which they are found are quite diverse. The α - and β -chains of horse methemoglobin, for

example, are nearly identical with those of human deoxyhemoglobin (Muirhead *et al.*, 1967), but the assembly into tetramers is somewhat different in the two cases, and the packing of these tetramers into their unit cells is very different indeed.

Situations like the above, in which a known molecular structure occurs in a variety of interesting crystallographic arrangements, are likely to arise for many large and important biological molecules. Much labor would be saved in these cases if the relevant crystal structures were determined starting from the known molecular structure, rather than *ab initio*. To piece together a structure in this way one must be able to find the orientation and location of each molecule in the unit cell. The problem of fixing the translations has been attacked by Nordman & Nakatsu (1963), Rossmann, Blow, Harding & Coller (1964), Tollin (1966) and Crowther & Blow (1967). The fundamental work on determining the orientations is that of Rossmann & Blow (1962). Tollin (1969) has combined these techniques to effect a complete protein structure determination.

* A portion of this work was presented in the Ph. D. thesis submitted (by EEL) to the Thomas C. Jenkins Department of Biophysics, The Johns Hopkins University.

In what follows a 'Patterson function' denotes the convolution of a structure with its centrosymmetric image, and may be periodic or aperiodic as the case warrants. A 'self-Patterson function' is one containing only intramolecular vectors, whereas a 'cross-Patterson function' contains only intermolecular vectors.

Let P_1 and P_2 be Patterson functions. Rossmann & Blow (1962) compare P_1 and P_2 as a function of their relative orientation. Their measure of agreement, R , is large whenever a self-Patterson function in P_1 lies parallel to an identical or similar function in P_2 . The Fourier coefficients of P_1 and P_2 are required for the calculation, which is done in reciprocal space for convenience. We discuss here a computationally swift version of the Rossmann-Blow approach which we have used to compare the Patterson function of an isolated molecule of sperm-whale metmyoglobin (SWMb) with the Patterson function of the D2 crystals of lamprey hemoglobin (Hendrickson, Love & Murray, 1968), and also for other experiments.

Nordman & Nakatsu (1963) have used a minimum function to compare directly the Patterson functions of a crystal and of a known, isolated molecule. More recently Nordman (1969) has used a modified minimum function to find the directions of the heme normal and of the axes of the α -helical sections in the SWMb molecule. Technical details are given by Nordman (1966) and Schilling (1968).

Zwick (1969), using a reciprocal space approach similar to ours, has also determined the direction of the heme normal and the helix axes. In addition he has found the orientation about the helix axis, and the position, of one of the helical segments.

Sarma (1969) at Oxford has investigated the triclinic form of lysozyme by forming the usual crystallographic R value between the squared transform of one molecule and the diffraction pattern from the crystal as a function of their relative orientation.

What happens to these methods when differences between the known molecule and the molecules in the crystal cannot be ignored? A trial structure can be obtained, as outlined above, by finding the orientation and position at which a known molecule fits best in the unit cell of a crystal. But some type of refinement is then required in order to achieve an accurate structure determination. For proteins only two possibilities suggest themselves: direct methods, which have not yet been fully tested on macromolecules, and methods using non-crystallographic symmetry (Main & Rossmann, 1966; Muirhead *et al.*, 1967; Maslen, 1968), which are not always applicable. Clearly the technique of assembling accurate crystal structures from the known structures of the constituent molecules is far from complete.

Methods

Rossmann & Blow (1962) studied a rotational correlation function, R , which can be written

$$R(\mathbf{C}) = \int_{-\infty}^{\infty} P_1(\mathbf{x})U(\mathbf{x})P_2(\mathbf{C}\mathbf{x})dV. \quad (1)$$

Here P_1 and P_2 are Patterson functions, \mathbf{C} is a variable rotation matrix, and U is a shape function having value one within a chosen volume (usually a sphere) and value zero outside. By Fourier transformation of the right-hand side they showed that, apart from constants of proportionality,

$$R(\mathbf{C}) = \sum_{\mathbf{p}} \sum_{\mathbf{h}} F_2^2(\mathbf{p})F_1^2(\mathbf{h})G(\mathbf{h}+\mathbf{h}'). \quad (2)$$

The summations extend over all values of \mathbf{h} and \mathbf{p} , the reciprocal lattice vectors of P_1 and P_2 ; F_1^2 and F_2^2 are the corresponding intensities. The non-integral reciprocal lattice vector \mathbf{h}' is given by $\tilde{\mathbf{C}}\mathbf{p}$, where $\tilde{\mathbf{C}}$ is the transpose of the matrix \mathbf{C} , and G is the Fourier transform of U .

In the Rossmann-Blow formulation P_1 and P_2 are periodic. As they point out, however, no operational limitation arises, since (2) applies to any Patterson function when properly 'crystallized'.

The sum over \mathbf{h} in (2) is, in fact, a convolution whose value is the Fourier transform of P_1U , which we term F_U^2 . We can therefore write that

$$R(\mathbf{C}) = \sum_{\mathbf{p}} F_2^2(\mathbf{p})F_U^2(\tilde{\mathbf{C}}\mathbf{p}), \quad (3)$$

in which the summation need extend only over one hemisphere of reciprocal space. In the special case in which P_1 is the Patterson function P_M of an isolated molecule M , the shape function U is no longer necessary since M itself is bounded. Letting F_M^2 represent the intensity transform of M , and rotating P_1 instead of P_2 , we have

$$R(\mathbf{C}) = \sum_{\mathbf{p}} F_M^2(\tilde{\mathbf{C}}\mathbf{p})F_2^2(\mathbf{p}), \quad (4)$$

which is the equation for the rotation function that we have used.

Considering the fast Fourier transform programs (Cooley & Tuckey, 1965; Cooley, Lewis & Welch, 1967) now available to evaluate the quantities F_U^2 or F_M^2 we believe that equation (3) or (4) will usually require less computation time than (2). The choice between direct and reciprocal space calculation, however, must be made individually for each problem. The integral in (1) will always be evaluated digitally; the relevant consideration is the number of sample points required for this evaluation *versus* the number of terms occurring in the summation in (3) or (4).

In order to improve clarity or speed of computation we have made a number of modifications in equation (4). As suggested by Rossmann & Blow we have generally omitted the low-order reflections when comparing proteins. This technique can substantially reduce the number of extraneous peaks in the rotation function, no doubt because these near-in reflections are strongly contaminated by scattering from the intermolecular mother liquor.

It is clear from (1) that, if the molecular Patterson function P_M overlaps the large values at adjacent origins of P_2 , strong but physically meaningless contributions will be made to R . When such overlap occurs removal of the origin in P_2 can improve the clarity of the results. We have particularly noted this effect (Lattman, 1969) in a comparison involving 6-azidopurine, a planar molecule whose length is several times greater than that of the b axis of the crystal in which it occurs (Glusker, van der Helm, Love, Minkin & Patterson, 1968). When the possibility of overlap is present we replace the values of F_2^2 in (4) by the Fourier coefficients of P_2 with its origin removed. For proteins we have used only reflections with Bragg spacings larger than 6\AA —'within the 6\AA sphere'. At this resolution we can employ the relation (Lipson & Cochran, 1966)

$$F'^2 = F_2^2 - \bar{F}_2^2, \quad (5)$$

where F_2^2 is the desired coefficient and \bar{F}_2^2 is the mean value of F_2^2 within the sphere of data used.

Again following Rossmann & Blow we have omitted terms in (4) for which the magnitude of F_2^2 (or F_2^2) is small, effecting a considerable reduction in computing time. When removal of the origin is not necessary, we have found that only 15–20% of the reflections are required to produce clear maps.

In order to assess the significance of peaks in R we have found it useful to compute Δ , the root-mean-square fluctuation of R , which is given by

$$\Delta^2 = \int [R(\theta) - \bar{R}]^2 dV. \quad (6)$$

Here \bar{R} is the mean value of R , and θ ($=\theta_1, \theta_2, \theta_3$) is the triple of Eulerian angles (Goldstein, 1959) defining C . We have found (Lattman, 1969) in all our trials that a peak whose height exceeds that of all other peaks by at least Δ corresponds to an actual alignment of the self-Patterson functions of interest. All our maps are scaled to a maximum value of 150 for ease in contouring. At this level typical values of Δ are between 20 and 25, while the minimum value of R is in the range -20 to 20 .

The computation of R is carried out on a grid in θ which spans the appropriate angular ranges (Tollin, Main & Rossmann, 1966). The grids we have used take fixed increments in each of the three Eulerian angles, producing uneven and inefficient sampling of R . A better sampling technique has been devised by Tollin & Cochran (1964). For every θ each of the vectors \mathbf{Ch} is computed in turn, and the value of $F_M^2(\mathbf{Ch})$ is obtained by three-dimensional, linear interpolation. The program when running on the IBM-7094 computer takes about 1 millisecond per reflection to compute one value of $R(\theta)$.

Most of our searches were done using the intensity transform of a molecule of SWMb (Bodo *et al.*, 1959; Kendrew, Dickerson, Strandberg, Hart, Davies, Phillips & Shore, 1960) which was computed by Fourier

transformation of its calculated electron density function ρ_W . We computed ρ_W using structure factors derived from the atomic positions given by Watson (1969). In these calculations an overall, isotropic temperature factor of 20\AA^2 was applied; structure factors were calculated to 6\AA resolution; ρ_W was evaluated at the nodes of a 2\AA cubic lattice; the intensity transform was similarly sampled, at intervals of $(1/128)\text{\AA}^{-1}$. The resultant transform has about 25,000 unique sample points within the 6\AA sphere. The computation of the transform directly by structure factor calculation alone would have taken too long. Problems of correct sampling encountered in this calculation are discussed by Goodman (1968).

Results

We have investigated the D2 crystals of cyanide-methemoglobin from the sea lamprey, *Petromyzon marinus*, with the rotation function, using SWMb as a test molecule. The resultant map of R displays only one significant peak. It is higher than any other by at least 1.2Δ . A portion of the map including the peak is shown in Fig. 1. In this calculation, for which the origin of the Patterson function was retained, we used about one quarter of the reflections within the 7\AA hemisphere, or some 200 data. Reflections with Bragg spacings larger than 12\AA were not used. We explored each Eulerian angle in the range 0 – 180 degrees, using 15 degree steps. A similar calculation in which the

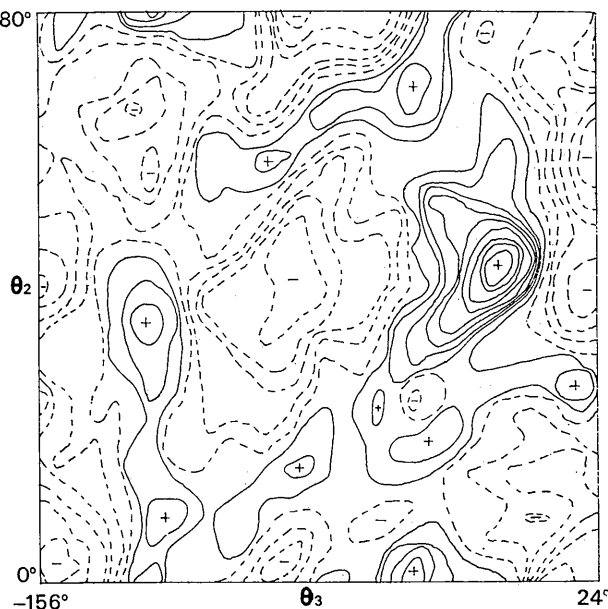


Fig. 1. Section through $R(\theta)$ for the SWMb/Lamprey hemoglobin comparison: the major peak in R is on the right, at $\theta_1 = -132$, $\theta_2 = -100$, $\theta_3 = -10^\circ$. In order to better display the peak, unconventional limits on θ_3 have been chosen. The space group of this rotation function is $P2_1ab$, retaining the order $\theta_1, \theta_2, \theta_3$. The unit cell is defined by $0 \leq \theta_1 < 2\pi$, $0 \leq \theta_2 < 2\pi$, $0 \leq \theta_3 < \pi$. This section was calculated on a 5° grid for clarity. Contours below the mean are hatched.

Patterson function origin was deleted did not yield significantly different results.

While this paper was being revised the structure of these D2 crystals was determined in this laboratory, and it is clear that the predictions of the rotation function are fulfilled: single chains of lamprey hemoglobin do resemble the SWMb molecule, and the orientation of these chains in the unit cell is as indicated by the position of the peak in R .

Lamprey hemoglobin D2 crystals belong to the space group $P2_12_12_1$ and have one molecule weighing about 18000 Daltons in the asymmetric unit. In this case the ratio of the numbers of cross- and self-Patterson functions is 3:1, compared with a 1:1 ratio encountered in various test problems involving monoclinic space groups. The sensitivity of the rotation function may be expected to decline as this ratio increases. We were pleased to see, however, that this decline was not noticeable in going from the 1:1 to the 3:1 ratio.

Of the various test problems studied with the rotation function (Lattman, 1969) only one was of concern. A comparison of a crystalline hemoglobin from the marine annelid, *Glycera dibranchiata*, with a test molecule of SWMb gave three false peaks essentially as high as the correct one. Yet the *Glycera* hemoglobin molecule is known to resemble SWMb (Padlan & Love, 1968). In addition, a control comparison of this crystalline *Glycera* hemoglobin with an isolated molecule of the same material did give a correct and unambiguous result. We have no convincing explanation for these observations.

Conclusions

We have developed a modification of the Rossmann-Blow rotation function in which the squared transform of a molecule is compared with the intensity set from a crystal. It can be rapidly evaluated. We have used it to show that molecules in the D2 crystals of lamprey hemoglobin are closely similar to the sperm-whale metmyoglobin molecule, and to find how they are oriented in the unit cell of this crystal. Using this method, we hope to investigate crystals forms of lamprey hemoglobin having polymeric asymmetric units.

We should like to thank Drs Helen Scouloudi, Herman Watson, Eduardo Padlan and Michael Rossmann for the use of their data. We have had helpful discussions with Drs David Blow, Michael Rossmann, Jon Herriott and Wayne Hendrickson.

One of us (EEL) has been supported by pre-doctoral fellowship GM-15,907 from the USPHS, and by a fellowship from USPHS grant GM-716. The project was supported by grant AM-2528 from the USPHS. Additional funds for computation came from a grant

from the Faculty of Arts and Sciences of the Johns Hopkins University, and from USPHS grant GM-716.

References

- BODO, G., DINTZIS, H. M., KENDREW, J. C. & WYCKOFF, H. W. (1959). *Proc. Roy. Soc. A* **253**, 70.
- COOLEY, J. W., LEWIS, P. A. K. & WELCH, P. D. (1967). *The Fast Fourier Transform and its Applications*. IBM research paper RC-1743, Yorktown Heights, N.Y.
- COOLEY, J. W. & TUCKEY, J. W. (1965). *Mathematics of Computation*, **19**, 297.
- CROWTHER, R. A. & BLOW, D. M. (1967). *Acta Cryst.* **23**, 544.
- CULLIS, A. F., MUIRHEAD, H., PERUTZ, M. F., ROSSMANN, M. G. & NORTH, A. C. T. (1962). *Proc. Roy. Soc. A* **265**, 161.
- GLUSKER, J. P., VAN DER HELM, D., LOVE, W. E., MINKIN, J. A. & PATTERSON, A. L. (1968). *Acta Cryst.* **B24**, 359.
- GOLDSTEIN, H. (1959). *Classical Mechanics*, p. 93. Reading: Addison-Wesley.
- GOODMAN, J. (1968). *Introduction to Fourier Optics*, p. 21. New York: McGraw-Hill.
- HENDRICKSON, W. A., LOVE, W. E. & MURRAY, G. (1968). *J. Mol. Biol.* **33**, 829.
- HUBER, R., FORMANEK, H. & EPP, O. (1968). *Naturwissenschaften*, **2**, 75.
- KENDREW, J. C., DICKERSON, R. E., STRANDBERG, B. E., HART, R. G., DAVIES, D. R., PHILLIPS, D. C. & SHORE, V. C. (1960). *Nature, Lond.* **185**, 422.
- LATTMAN, E. (1969). Thesis, Johns Hopkins University.
- LIPSON, H. & COCHRAN, W. (1966). *The Determination of Crystal Structures*, p. 170. London: Bell.
- MAIN, P. & ROSSMANN, M. G. (1966). *Acta Cryst.* **21**, 67.
- MASLEN, E. N. (1968). *Acta Cryst.* **B24**, 1165.
- MUIRHEAD, H., COX, J. M., MAZZARELLA, L. & PERUTZ, M. F. (1967). *J. Mol. Biol.* **28**, 117.
- NORDMAN, C. (1966). *Trans. Amer. Cryst. Assoc.* **2**, 29.
- NORDMAN, C. (1969). Personal communication.
- NORDMAN, C. E. & NAKATSU, K. (1963). *J. Amer. Chem. Soc.* **85**, 353.
- PADLAN, E. A. & LOVE, W. E. (1968). *Nature, Lond.* **220**, 376.
- ROSSMANN, M. G. & BLOW, D. M. (1962). *Acta Cryst.* **15**, 24.
- ROSSMANN, M. G., BLOW, D. M., HARDING, M. M. & COLLIER, E. (1964). *Acta Cryst.* **17**, 338.
- SARMA, V. R. (1969). Personal communication.
- SCHILLING, J. C. (1968). Thesis, University of Michigan.
- SCOULOUDI, H. (1969). *J. Mol. Biol.* **40**, 353.
- TOLLIN, P. (1966). *Acta Cryst.* **21**, 613.
- TOLLIN, P. (1969). *J. Mol. Biol.* **45**, 481.
- TOLLIN, P. & COCHRAN, W. (1964). *Acta Cryst.* **17**, 1322.
- TOLLIN, P., MAIN, P. & ROSSMANN, M. G. (1966). *Acta Cryst.* **20**, 404.
- WATSON, H. C. (1969). In *Progress in Stereochemistry*, **4**, Ed. W. KLYNE & P. DE LA MARE. In the press. New York City: Academic Press.
- ZWICK, M. (1969). *Abstr. Amer. Cryst. Assoc. Winter Meeting*, p. 74.