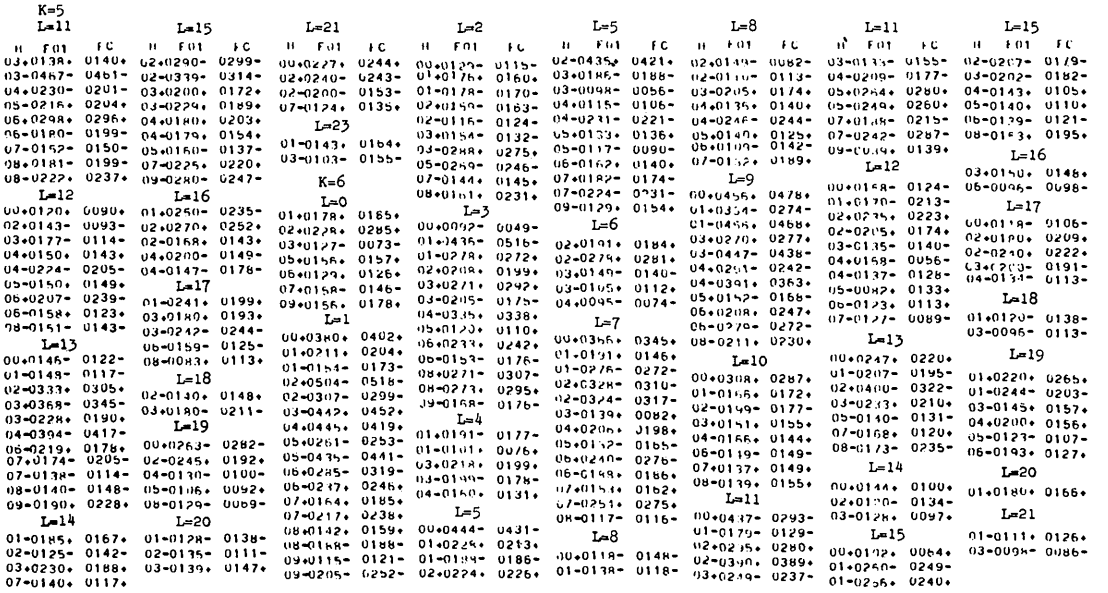
Table 8 (*cont.*)



Teller effect will perhaps become clearer after further investigation of the influence of the alkyl chains. In order to throw some light on this question we are studying the structure of bis(*N*-n-propylsalicylaldiminato)copper(II) and of bis(*N*-n-heptylsalicylaldiminato)nickel(II).

References

CHARLES, R. G. (1957). *J. Org. Chem.* **22**, 677.

CHEESEMAN, T. P., HALL, D. & WATERS, T. N. (1963). *Proc. Chem. Soc.* p. 379.

FORSYTH, J. B. & WELLS, M. (1959). *Acta Cryst.* **12**, 412.

FRASSON, E. & PANATTONI, C. (1962). *Proc.* VII *Intern. Conf. Coordination Chemistry*, Stockholm. p. 89

JARSKI, M. A. & LINGAFELTER, E. C. (1964). *Acta Cryst.* **17**, 1109.

LINGAFELTER, E. C., SIMMONS, G. L., MOROSIN, B., SCHERINGER, C. & FREIBURG, C. (1961). *Acta Cryst.* **14**, 1222.

ORIOLI, P. L., LINGAFELTER, E. C. & BROWN, B. W. (1964). *Acta Cryst.* **17**, 1113.

PANATTONI, C. & FRASSON, E. (1962). *Ric. Sci.* **32**, 151.

PANATTONI, C., GRUBISSICH, C. & FRASSON, E. (1965). *Ric. Sci.*, II-A, **35**, 995.

PAULING, L. (1960). *The Nature of the Chemical Bond.* 3d. edition. Ithaca: Cornell Univ. Press.

SACCONI, L., CIAMPOLINI, M., MAGGIO, F. & DEL RE, G. (1960). *J. Amer. Chem. Soc.* **82**, 815.

SCATTURIN, W., PANATTONI, C., BARDI, R. & BEZZI, S. (1959). *Gazz. chim. ital.* **89**, 2142.

STARIKOVA, Z. A. (1966). *Acta Cryst.* **21**, A 154.

WEI, L., STOGSDILL, R. M. & LINGAFELTER, E. C. (1964). *Acta Cryst.* **17**, 1058.

Acta Cryst. (1967). **23**, 544

A Method of Positioning a Known Molecule in an Unknown Crystal Structure

BY R. A. CROWTHER AND D. M. BLOW

Medical Research Council Laboratory of Molecular Biology, Hills Road, Cambridge, England

(*Received* 19 *April* 1967)

A function is proposed for determining the position of a known molecule of known orientation relative to a crystallographic symmetry element in an unknown crystal structure. A Fourier-series summation with appropriate coefficients is used to express the correlation between the observed Patterson function of the crystal and the set of cross-Patterson vectors of a model structure. The point of maximum correlation gives the value of the intermolecular vector between molecules related by the chosen symmetry element. The function has been tested on sperm whale myoglobin.

Introduction

A number of methods exist for determining the orientation of a known molecule in an unknown crystal structure (Rossmann & Blow, 1962; Nordman & Nakatsu, 1963; Tollin & Cochran, 1964). Several methods have also been used to solve the subsequent translational problem of positioning the molecule rela-

tive to the chosen origin of the crystal (Nordman & Nakatsu, 1963; Tollin & Cochran, 1964; Tollin, 1966). We wish to present a new method of solving the translational problem. The proposed function is simple to compute and, we believe, particularly well suited to the analysis of complex molecules. We expect to be able to apply it to the study of crystalline proteins in cases where a closely related protein structure is known (Scouloudi, 1960).

Consider the problem of positioning a known molecule relative to a particular crystallographic symmetry element. The term 'known molecule' here implies that not only do we know the atomic coordinates relative to some local origin O, fixed in the molecule, but also that the molecule is in the same orientation as one of the molecules in the unknown crystal structure. We wish to determine the position of the local origin O with respect to the chosen crystallographic symmetry element. If the known molecule is placed at an arbitrary position in the unit cell, the position of the symmetry related molecule is then fixed and it is possible to calculate the set of Patterson vectors from the known molecule to the symmetry related molecule. As the position of the known molecule varies, the relative configuration of this set of Patterson vectors is unchanged but the set moves bodily to a position characterized by the vector joining the local origins of the known molecule and the symmetry related molecule. This set of vectors will be referred to as the cross-Patterson of the model structure and is to be regarded as a function of two variables, namely position in Patterson space and also position of the local origin of the known molecule in the model structure.

A well known method of finding the position of the local origin O is to move this set of cross-Patterson vectors over the observed Patterson function of the crystal, using a minimum function or similar measure of fit to determine the correct solution (see Buerger, 1959). This method is most powerful when there are few atoms in the molecule. If there are very many atoms it would be natural to sum up the observed Patterson density at every point corresponding to an expected cross-Patterson peak and to use this total as a measure of fit for a particular position of O. This is equivalent to the convolution of the observed Patterson function with the group of cross-Patterson vectors of the model structure. This convolution may be achieved by multiplying the Fourier coefficients of the two functions, and performing a Fourier summation with the coefficients so obtained. The resulting function should have a prominent peak corresponding to the vector between the two origins. In the next section we give a formal derivation.

Derivation of the translation function

If an atom in the reference molecule has crystallographic coordinates $\mathbf{x}$, then let the corresponding atom in the symmetry related molecule have crystallographic coordinates $(\mathbf{Ax}+\mathbf{d})$ (Fig. 1). The set of cross-Patterson vectors from the reference molecule to the symmetry related molecule may be written as

$$P_{01}(\mathbf{u})=\int_V \varrho_0(\mathbf{x})\varrho_1(\mathbf{x}+\mathbf{u})d\mathbf{x} \qquad (1)$$

where ϱ_0, ϱ_1 represent the electron densities within the reference molecule and the symmetry related molecule respectively, and the integral is taken over the whole unit cell volume, V. If the local origin of the known molecule is at $\mathbf{s}$, we may express (1) in terms of the electron density, ϱ_M, of the known molecule relative to the local origin O.

$$P_{01}(\mathbf{u},\mathbf{s})=\int_V \varrho_M(\mathbf{x}-\mathbf{s})\varrho_M[\mathbf{A}^{-1}(\mathbf{x}+\mathbf{u}-\mathbf{As}-\mathbf{d})]d\mathbf{x}\,.$$

Expanding ϱ_M as a Fourier series in terms of the structure factors, F_M, of the known molecule calculated relative to the local origin O, we have

$$\begin{aligned}P_{01}(\mathbf{u},\mathbf{s}) &= \int_V \sum_{\mathbf{h}} F_M(\mathbf{h}) \exp[-2\pi i\mathbf{h}.(\mathbf{x}-\mathbf{s})] \\ &\quad \times \sum_{\mathbf{p}} F_M(\mathbf{p}) \exp[-2\pi i\mathbf{p}\mathbf{A}^{-1}(\mathbf{x}+\mathbf{u}-\mathbf{As}-\mathbf{d})]d\mathbf{x} \\ &= \sum_{\mathbf{h}}\sum_{\mathbf{p}} F_M(\mathbf{h})F_M(\mathbf{p}) \exp[2\pi i(\mathbf{h}.\mathbf{s}+\mathbf{p}\mathbf{A}^{-1}(\mathbf{As}+\mathbf{d}))] \\ &\quad \exp(-2\pi i\mathbf{p}\mathbf{A}^{-1}\mathbf{u})\int_V \exp[-2\pi i(\mathbf{h}+\mathbf{p}\mathbf{A}^{-1})\mathbf{x}]d\mathbf{x}\,.\end{aligned}$$

Since $\mathbf{A}$ is a crystallographic symmetry operator the integral vanishes unless $\mathbf{h}+\mathbf{p}\mathbf{A}^{-1}=0$, when it takes the value unity. Therefore,

$$\begin{aligned}P_{01}(\mathbf{u},\mathbf{s}) &= \sum_{\mathbf{h}} F_M(\mathbf{h})F_M(-\mathbf{hA}) \\ &\quad \exp[2\pi i\mathbf{h}.(\mathbf{s}-\mathbf{As}-\mathbf{d})] \exp(2\pi i\mathbf{h}.\mathbf{u})\,.\end{aligned}$$

Writing the intermolecular vector $\mathbf{t}=-\mathbf{s}+\mathbf{As}+\mathbf{d}$ (Fig. 1) and using the Friedel relation, this becomes

$$\begin{aligned}P_{01}(\mathbf{u},\mathbf{t}) &= \sum_{\mathbf{h}} F_M(\mathbf{h})F_M^*(\mathbf{hA}) \\ &\quad \exp(-2\pi i\mathbf{h}.\mathbf{t}) \exp(2\pi i\mathbf{h}.\mathbf{u})\,. \qquad (2)\end{aligned}$$

We now define the translation function $T(\mathbf{t})$ by the convolution

$$T(\mathbf{t})=\int_V P_{01}(\mathbf{u},\mathbf{t})P(\mathbf{u})d\mathbf{u}\,, \qquad (3)$$

where $P(\mathbf{u})$ represents the observed Patterson function of the crystal. When $\mathbf{t}$ becomes equal to the 'true' intermolecular vector $\mathbf{t}_0$, the computed cross-Patterson vectors P_{01} fit correctly to the observed Patterson function P, and $T(\mathbf{t}_0)$ will have a large positive value.

Expanding the observed Patterson function $P(\mathbf{u})$ as a Fourier series and using the expression (2) we may write (3) as

$$\begin{aligned}T(\mathbf{t}) &= \int_V \sum_{\mathbf{h}} F_M(\mathbf{h})F_M^*(\mathbf{hA}) \exp(-2\pi i\mathbf{h}.\mathbf{t}) \\ &\quad \exp(2\pi i\mathbf{h}.\mathbf{u}) \sum_{\mathbf{p}} |F_{\mathrm{obs}}(\mathbf{p})|^2 \exp(-2\pi i\mathbf{p}.\mathbf{u})d\mathbf{u}\,.\end{aligned}$$

The integral vanishes unless $\mathbf{h}-\mathbf{p}=0$, so that we have finally

$$T(\mathbf{t})=\sum_{\mathbf{h}} |F_{\rm obs}(\mathbf{h})|^2 F_M(\mathbf{h})F_M^*(\mathbf{hA}) \exp(-2\pi i \mathbf{h}.\mathbf{t}) .$$

We have thus expressed the correlation between a set of cross-Patterson vectors for a model structure and the observed Patterson function in terms of a Fourier series with coefficients which are simple to compute. The translation function can then be evaluated by means of a standard Fourier summation program.

Modified translation functions

The observed Patterson function of the crystal contains both intramolecular and intermolecular Patterson vectors. Since we are interested in fitting the intermolecular vectors, the intramolecular vectors only serve to increase the background noise in the summation. However, since the molecular structure is known, it is possible to remove the intramolecular vectors from the observed Patterson function, provided the observed intensities can be put on an appropriate scale. Suppose that there are n molecules in the unit cell and that the ith molecule is related to the known molecule by a symmetry operator whose rotational part is represented by matrix $\mathbf{A}_i$, where $\mathbf{A}_0$, the matrix relating the known molecule to itself, is the identity matrix. Then the modified translation function, which fits the set of intermolecular vectors between the known molecule and the chosen symmetry related molecule to the complete set of intermolecular vectors in the observed Patterson function is given by

$$T_1(\mathbf{t})=\sum_{\mathbf{h}} \left(|F_{\rm obs}(\mathbf{h})|^2 - \sum_{i=0}^{n-1} |F_M(\mathbf{hA}_i)|^2\right) \times F_M(\mathbf{h})F_M^*(\mathbf{hA}) \exp(-2\pi i \mathbf{h}\,.\,\mathbf{t}) ,$$

where $F_{\rm obs}$ and F_M are assumed to be on an absolute scale.

Unless the molecule has a symmetry of its own, $F_M(\mathbf{h})$ has no symmetry, and in general the functions $T(\mathbf{t}), T_1(\mathbf{t})$ will have only $P1$ symmetry. However, because of Friedel's Law, some special cases arise. Let us write

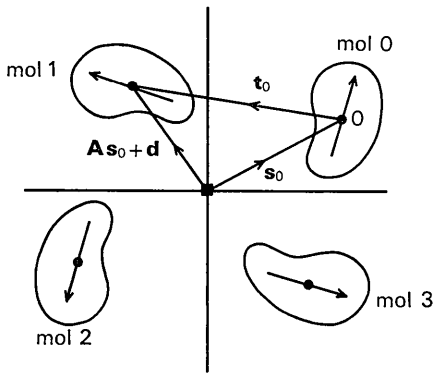
$$F_M(\mathbf{h})F_M^*(\mathbf{hA})=F_M(\mathbf{h})F_M(-\mathbf{hA}) .$$

If $\mathbf{A}$ is a twofold rotation, $-\mathbf{A}$ is a reflexion in a perpendicular mirror plane, and since Friedel's Law also enforces this symmetry on $|F_{\rm obs}(\mathbf{h})|$, the translation functions have Pm symmetry. Conversely if $\mathbf{A}$ is a mirror operation, the translation functions have $P2$ symmetry. In each of these cases, the translation vector $\mathbf{t}_0$ must always lie in a special position in a plane or on an axis.

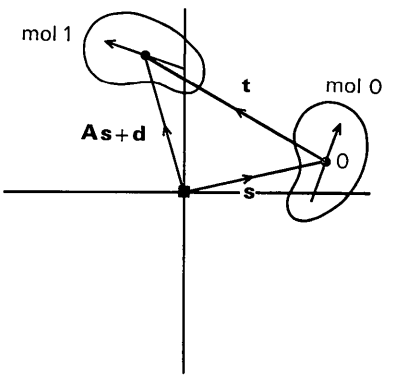
The expressions given so far represent the fitting of one particular set of cross-Patterson vectors of a model structure containing two molecules to the complete set of intermolecular vectors in the observed Patterson function. Which intermolecular vector is found depends on which two molecules were used in the model structure. It is possible, however, to modify the expression for the translation function in such a way that it has higher symmetry and contains peaks corresponding to all possible intermolecular vectors in the unknown structure.

To do this we must use a model structure which contains the same number of molecules as the unknown structure. Using the above notation the set of cross-Patterson vectors from molecule i to molecule j may be written, by comparison with equation (1) above, as

$$P_{ij}(\mathbf{u},\mathbf{t}_{ij})=\sum_{\mathbf{h}} F_M(\mathbf{hA}_i)F_M^*(\mathbf{hA}_j) \exp(-2\pi i \mathbf{h}.\mathbf{t}_{ij}) \exp(2\pi i \mathbf{h}.\mathbf{u}) ,$$



(a)



(b)

Fig. 1. (*a*) Unknown structure containing four molecules in space group $P4$. In this case $\mathbf{A}=\begin{pmatrix}0 & -1 & 0\\ 1 & 0 & 0\\ 0 & 0 & 1\end{pmatrix}$ $\mathbf{d}=(0,0,0)$.

(*b*) Model structure containing two molecules in correct orientation relative to the unknown structure and related by a fourfold axis, but at an otherwise arbitrary position in the unit cell. As vector $\mathbf{s}$ varies, the correlation between the set of Patterson vectors from molecule 0 to molecule 1 in the model structure and the observed Patterson function of the unknown structure takes its largest value when $\mathbf{s}=\mathbf{s}_0$. The peak in the translation function representing the required intermolecular vector occurs at the point $\mathbf{t}_0$.

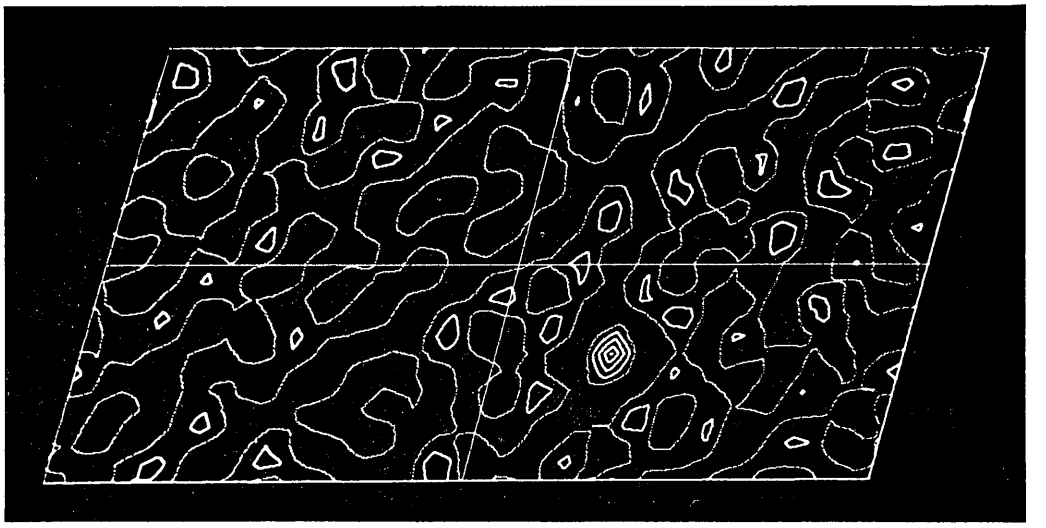
where $\mathbf{t}_{ij}$ is now the intermolecular vector between molecule i and molecule j. If we sum over all possible pairs (i,j) and convolute the resulting expression with the observed Patterson function we obtain

$$\sum_{\mathbf{h}} |F_{\text{obs}}(\mathbf{h})|^2 [\sum_{\substack{i=0 \\ i\neq j}}^{n-1} \sum_{j=0}^{n-1} F_M(\mathbf{hA}_i) F_M^*(\mathbf{hA}_j) \exp(-2\pi i \mathbf{h}.\mathbf{t}_{ij})] \ .$$

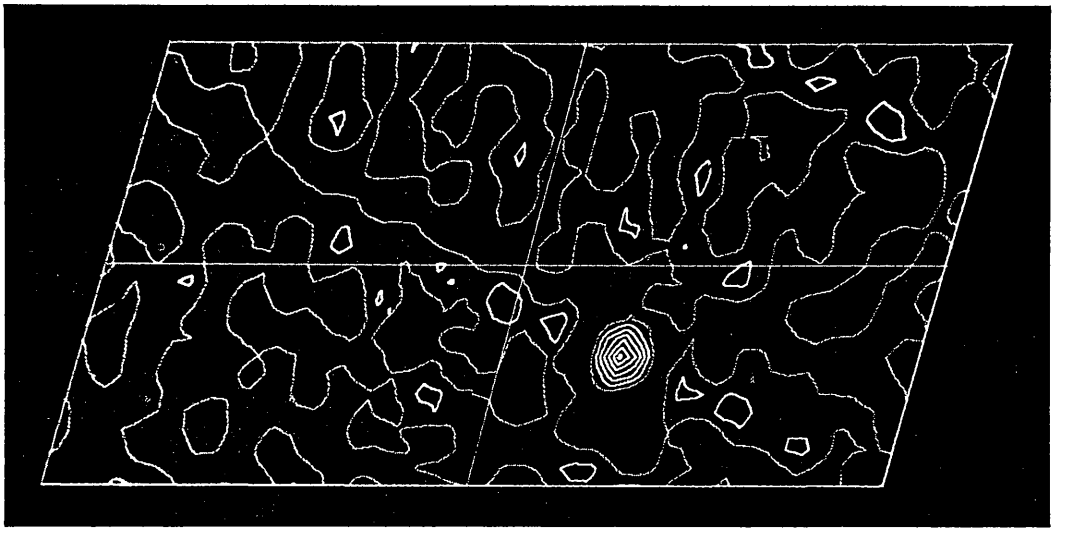
Since the exponent appears within the inner summations, this function cannot be computed by a single Fourier summation. However, a Fourier summation of the form

$$T_2(\mathbf{t}) = \sum_{\mathbf{h}} |F_{\text{obs}}(\mathbf{h})|^2 [\sum_{\substack{i=0 \\ i\neq j}}^{n-1} \sum_{j=0}^{n-1} F_M(\mathbf{hA}_i) F_M^*(\mathbf{hA}_j)] \exp(-2\pi i \mathbf{h}.\mathbf{t})$$

will have peaks for all possible intermolecular vectors in the unknown structure. The spatial arrangement and relative weights of the peaks will be the same as the peaks in the Patterson function of the point group of the molecules within the unit cell of the unknown structure, but with the origin peak removed. As before, the intramolecular vectors may be removed from the observed Patterson function before convoluting it with



(*a*)



(*b*)

Fig. 2. (*a*) Two-dimensional translation function $T_1(t_x,t_z)$ for sperm whale myoglobin (case 2, Table 1). (*b*) Section $y=\frac{1}{2}$ from three-dimensional translation function $T_1(t_x,\frac{1}{2},t_z)$ for sperm whale myoglobin (case 3, Table 1). In each case t_x is horizontal. Negative contours have been suppressed and the zero contour has been plotted half-weight. The expected peak, for the particular choice of local origin made, should occur at (0·645, 0·708).

the model Patterson function. This function $T_2(\mathbf{t})$ is no better than $T(\mathbf{t})$ from the point of view of the ratio of peak and background, though because of its higher symmetry it may be easier to compute.

Application

The method has been tested by applying it to sperm whale myoglobin. This protein has a molecular weight of approximately 17000 and crystallizes in space group $P2_1$ with two molecules in the unit cell. The expressions for the translation function in this case become

$$T(t_x,t_y,t_z)=\sum_h\sum_k\sum_l |F_{\rm obs}(hkl)|^2F_M(hkl)F_M(h\bar{k}l)$$
$$\exp[-2\pi i(ht_x+kt_y+lt_z)]$$

$$T_1(t_x,t_y,t_z)=\sum_h\sum_k\sum_l (|F_{\rm obs}(hkl)|^2-|F_M(hkl)|^2$$
$$-|F_M(h\bar{k}l)|^2)F_M(hkl)F_M(h\bar{k}l)$$
$$\exp[-2\pi i(ht_x+kt_y+lt_z)]\,.$$

Since the relative y coordinate of the two molecules is known to be $\frac{1}{2}$, it is only necessary to compute the section $y=\frac{1}{2}$ in the Fourier summation. Alternatively it should be possible to determine the relative x and z coordinates by means of a two-dimensional summation. The translation function in projection becomes:

$$T(t_x,t_z)=\sum_h\sum_l |F_{\rm obs}(h0l)|^2F_M^2(h0l)$$
$$\exp[-2\pi i(ht_x+lt_z)]$$

$$T_1(t_x,t_z)=\sum_h\sum_l (|F_{\rm obs}(h0l)|^2-2|F_M(h0l)|^2)F_M^2(h0l)$$
$$\exp[-2\pi i(ht_x+lt_z)]\,.$$

The molecular structure factors were calculated from atomic coordinates kindly supplied by Dr H.C.Watson and Dr J.C.Kendrew. For the particular choice of local origin used in calculating the molecular structure factors, the required peak should occur at $t_x=0{\cdot}645$, $t_z=0{\cdot}708$. The results are summarized in Table 1. Cases 1 and 2 relate to calculations in projection including only $h0l$ reflexions with spacings between 8 and 4 Å (approximately 180 terms). Case 3 relates to section $y=\frac{1}{2}$ from a three-dimensional summation, again including reflexions between 8 and 4 Å (approximately 1000 terms). Fig.2 shows the translation function as calculated in cases 2 and 3.

Table 1. *Summary of results for sperm whale myoglobin*

Case	Function	Ratio of expected to next highest peak
1	T (projection)	1·60
2	T_1 (projection)	2·14
3	T_1 (section)	3·28

Table 1 shows that the results are improved by using the function T_1 rather than T. Reflexions corresponding to spacings greater than 8 Å were omitted, because the observed structure amplitudes include large contributions to low order terms from the mother liquor, which fills the spaces between the protein molecules in the crystal. The shapes of these regions do not obey the required translation relations.

It is extremely encouraging that the method gives such clear cut results even when working in projection, when the number of terms included in the summation is comparatively small. The number of terms cannot be reduced much further: when the summation is restricted to $h0l$ reflexions between 8 and 5 Å (approximately 100 terms), no significant peak is obtained.

We are grateful to Dr H.C.Watson and Dr J.C. Kendrew for providing atomic coordinates for sperm whale myoglobin. We thank Professor Gill and the Manager of the Computer Centre, Imperial College of Science and Technology, for making computing facilities available. A contour-plotting program written by Mr T.H.Gossling was used in preparing Fig.2. One of us (R.A.C.) is a holder of a Medical Research Council scholarship.

References

BUERGER, M. J. (1959). *Vector Space*, chapters 10 and 11. New York: John Wiley.

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

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

SCOULOUDI, H. (1960). *Proc. Roy. Soc.* A, **258**, 181.

TOLLIN, P. (1966). *Acta Cryst.* **21**, 613.

TOLLIN, P. & COCHRAN, W. (1964). *Acta Cryst.* **17**, 1322.