

The Detection of Sub-Units Within the Crystallographic Asymmetric Unit

BY MICHAEL G. ROSSMANN AND D. M. BLOW

M. R. C. Unit for Molecular Biology, Cavendish Laboratory, Cambridge, England

(Received 27 February 1961)

The number of structurally identical units within one unit cell often exceeds the number of general positions. The angular relationships between any two units, not related by space-group symmetry, can be found by rotating the Patterson function until the rotated and original Patterson functions are brought into maximum coincidence. For such a rotation, the rotation function

$$R = \sum_{\mathbf{h}} [F_{\mathbf{h}}]^2 \left\{ \sum_{\mathbf{p}} |F_{\mathbf{p}}|^2 G \right\}$$

has a maximum value. G is an interference function which has large values only when the point \mathbf{p} in reciprocal space is brought close to \mathbf{h} by the rotation.

Application of the R function to horse haemoglobin gives a dominant peak that corresponds accurately to the relative orientation of the α and β chains.

Evidence is accumulating that many of the larger protein molecules are made up of identical, or closely similar sub-units. The reasons for expecting this for the protein part of virus structures were set out by Crick & Watson (1956), and the prediction has been amply confirmed in crystallographic studies of spherical viruses (Caspar, 1956; Finch & Klug, 1959; Klug & Finch, 1960; Magdoff, 1960) and of one rod-shaped virus (Watson, 1954; Franklin & Holmes, 1958). Harrison (1959) has given crystallographic evidence that the large protein, ferritin, is made up of a number of sub-units. By far the most remarkable example, however, is haemoglobin (Perutz *et al.*, 1960). Although this protein has four polypeptide chains, four times the molecular weight and exactly four times the iron content of the related protein myoglobin, chemical evidence showed that there were two distinct kinds of polypeptide chain and that their composition was quite different from that of myoglobin. It was therefore a surprise, (strongly supporting the view that the spatial configuration of a protein is important) to find that the haemoglobin molecule consisted of four similar units, each very like the myoglobin molecule as determined by Kendrew and his collaborators (Bodo *et al.* (1959)), and arranged in a roughly tetrahedral manner. It may be noted that in every case referred to above, the sub-unit has a molecular weight of the order $2 \cdot 10^4$.

While the viruses represent examples of exact but non-crystallographic symmetry of the independent particles (e.g. five-fold rotation symmetry and non-integral screw operations), the haemoglobin structure may be said to have partial, approximate symmetry. The symmetry is approximate as the two chains are chemically distinct, and to a small extent they have different configuration. Moreover the operation which superimposes one chain on to another, does not necessarily superimpose the other back on to the first, so that this symmetry operation, which is only satisfied by a part of the molecule, is called partial.

In this paper we describe how we have detected the existence of this partial, approximate symmetry from a knowledge of the intensities alone. The effect of non-crystallographic symmetry, whether partial or total, results in decreasing the size of the structure to be determined, while the number of observable intensities remains the same. This 'redundancy' in information might be used to help solve a structure. Also, the ideas presented here are as applicable to finding the relationship between similar molecules in different crystal lattices, as they are to finding the relative orientation of molecules (or sub-units within a molecule) in the same crystal lattice.

1. The rotation of Patterson syntheses

Consider a structure of two identical units which are in different orientations. The Patterson function of such a structure consists of three parts. There will be the self-Patterson vectors of one unit, being the set of interatomic vectors which can be formed within that unit, with appropriate weights. The set of self-Patterson vectors of the other unit will be identical, but they will be rotated from the first due to the different orientation. Finally, there will be the cross-Patterson vectors, or set of interatomic vectors which can be formed from one unit to another. The self-Patterson vectors of the two units will all lie in a volume extending from the origin by the overall dimensions of the units. Some or all of the cross-Patterson vectors will lie outside this volume.

Suppose the Patterson function is now superposed on a rotated version of itself. There will be no particular agreement except when one set of self-Patterson vectors of one unit has the same orientation as the self-Patterson vectors from the other unit. In this position, we would expect a maximum of agreement or 'overlap' between the two.

These ideas have often been used by workers

looking for evidence about the orientation of symmetrical groups like benzene rings in simple structures. Perutz *et al.* (1955) used similar considerations for predicting a structural resemblance between reduced human haemoglobin and horse methaemoglobin.

In § 2 an expression will be developed which allows the overlap to be calculated directly from the observed intensities, for a given rotation. In § 3 the physical significance of the expression obtained will be considered.

2. The rotation function

Consider a point whose position \mathbf{r} may be specified relative to three crystallographic axes $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$ by $\mathbf{r} = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + x_3\mathbf{a}_3$. The position relative to another set of axes $\mathbf{a}'_1, \mathbf{a}'_2, \mathbf{a}'_3$ is given by

$$\begin{aligned} x'_1 &= c_{11}x_1 + c_{12}x_2 + c_{13}x_3 + d_1 \\ x'_2 &= c_{21}x_1 + c_{22}x_2 + c_{23}x_3 + d_2 \\ x'_3 &= c_{31}x_1 + c_{32}x_2 + c_{33}x_3 + d_3. \end{aligned} \quad (1)$$

Alternatively we can consider x'_1, x'_2, x'_3 as the coordinates of the point in the same axes, after it has been rotated and translated. For simplicity and brevity, matrix notation is preferable. All three of the above equations may then be written

$$\mathbf{x}' = [\mathbf{C}]\mathbf{x} + \mathbf{d}.$$

The d_i ($i = 1, 2, 3$) form a vector \mathbf{d} which represents the translation between the origins of the two systems, while the c_{ij} form a matrix $[\mathbf{C}]$ representing rotation about the origin of the unprimed system. In the application considered here, the translation \mathbf{d} will always be zero, but it will be retained for generality.

Now consider any function $\varrho(x_1, x_2, x_3)$ periodic in a cell defined by $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$. We define the *overlap*, R , of this function with a rotated and translated version of ϱ , within some volume U as

$$R = \iiint_U \varrho(x_1, x_2, x_3) \varrho(x'_1, x'_2, x'_3) dx_1 dx_2 dx_3. \quad (2)$$

It is clear that R will take large values if the transformation relating the primed and unprimed systems has the property that $\varrho(x_1, x_2, x_3)$ tends to be equal to $\varrho(x'_1, x'_2, x'_3)$ within the volume U .

Since ϱ is a periodic function it may be expanded in a Fourier series:

$$\varrho(x_1, x_2, x_3) = (1/V) \sum_{h_1} \sum_{h_2} \sum_{h_3} |F(h_1, h_2, h_3)| \times \exp \{i\alpha(h_1, h_2, h_3)\} \exp \{-2\pi i(h_1x_1 + h_2x_2 + h_3x_3)\}$$

or more briefly

$$\varrho(\mathbf{x}) = (1/V) \sum_{\mathbf{h}} |F(\mathbf{h})| \exp [i(\alpha_{\mathbf{h}} - \varphi_{\mathbf{h}, \mathbf{x}})],$$

where

$$\varphi_{\mathbf{h}, \mathbf{x}} = 2\pi \sum_{i=1}^3 h_i x_i.$$

Similarly we may write

$$\varrho(\mathbf{x}') = (1/V) \sum_{\mathbf{p}} |F(\mathbf{p})| \exp [i(\alpha_{\mathbf{p}} - \varphi_{\mathbf{p}, \mathbf{x}'})],$$

where the Fourier coefficient with indices (p_1, p_2, p_3) is written as $|F(\mathbf{p})| \exp [i\alpha_{\mathbf{p}}]$.

Since only the exponents $\varphi_{\mathbf{h}, \mathbf{x}}, \varphi_{\mathbf{p}, \mathbf{x}'}$ depend upon \mathbf{x} , the overlap may thus be written

$$R = (1/V^2) \sum_{\mathbf{h}} \sum_{\mathbf{p}} \left[|F_{\mathbf{h}}| |F_{\mathbf{p}}| \exp [i(\alpha_{\mathbf{h}} + \alpha_{\mathbf{p}})] \times \left\{ \iiint_U \exp [-i(\varphi_{\mathbf{h}, \mathbf{x}} + \varphi_{\mathbf{p}, \mathbf{x}'})] dx_1 dx_2 dx_3 \right\} \right].$$

Writing the exponents more fully we have

$$\begin{aligned} (\varphi_{\mathbf{h}, \mathbf{x}} + \varphi_{\mathbf{p}, \mathbf{x}'}) &= 2\pi \left(\sum_i h_i x_i + \sum_i p_i x'_i \right) \\ &= 2\pi \sum_i (h_i + h'_i) x_i + 2\pi \sum_i p_i d_i, \end{aligned}$$

where

$$\begin{aligned} h'_1 &= c_{11}p_1 + c_{21}p_2 + c_{31}p_3 \\ h'_2 &= c_{12}p_1 + c_{22}p_2 + c_{32}p_3 \\ h'_3 &= c_{13}p_1 + c_{23}p_2 + c_{33}p_3 \end{aligned} \quad (3)$$

or more briefly

$$\mathbf{h}' = [\tilde{\mathbf{C}}]\mathbf{p},$$

where $[\tilde{\mathbf{C}}]$ is the transpose of $[\mathbf{C}]$, obtained by interchanging its rows and columns.

Substituting back in the expression for R , we now have

$$R = (1/V^2) \sum_{\mathbf{h}} \sum_{\mathbf{p}} \left[|F_{\mathbf{h}}| |F_{\mathbf{p}}| \exp [i(\alpha_{\mathbf{h}} + \alpha_{\mathbf{p}} - 2\pi \sum_i p_i d_i)] \times \left\{ \iiint_U \exp [-2\pi i \sum_i (h_i + h'_i) x_i] dx_1 dx_2 dx_3 \right\} \right].$$

The bracketed integral is an interference function familiar in crystallography (see especially Patterson, 1939). We will write its solution as

$$(U/V) \times G_{\mathbf{h}, \mathbf{h}'} \exp [i\Omega_{\mathbf{h}, \mathbf{h}'}],$$

so that we have, finally

$$R = (U/V^3) \sum_{\mathbf{h}} \sum_{\mathbf{p}} |F_{\mathbf{h}}| |F_{\mathbf{p}}| G_{\mathbf{h}, \mathbf{h}'} \times \exp [i(\alpha_{\mathbf{h}} + \alpha_{\mathbf{p}} - 2\pi \sum_i p_i d_i + \Omega_{\mathbf{h}, \mathbf{h}'})]. \quad (4)$$

We quote as examples the expressions for G and Ω when the integral is bounded by the faces of a parallelepiped between $A_i^- \leq x_i \leq A_i^+$; and by the surface of a sphere radius r with centre at $(A_i^+ + A_i^-)/2$.

In both cases

$$\Omega_{\mathbf{h}, \mathbf{h}'} = \pi \sum_{i=1}^3 (h_i + h'_i) (A_i^- + A_i^+).$$

For the parallelepiped

$$G_{\mathbf{h}, \mathbf{h}'} = \prod_{i=1}^3 \frac{\sin \pi(h_i + h'_i)(A_i^+ - A_i^-)}{\pi(h_i + h'_i)(A_i^+ - A_i^-)} \quad (5)$$

and for the sphere

$$G_{\mathbf{h}, \mathbf{h}'} = \frac{3(\sin 2\pi Hr - 2\pi Hr \cos 2\pi Hr)}{(2\pi Hr)^3}$$

where H is the distance of the point

$$((h_1 + h_1'), (h_2 + h_2'), (h_3 + h_3'))$$

from the origin in reciprocal space.

In the application of R with which we are concerned, the above, generalized, form for R may be simplified. Since it is desirable to have a form dependent only on the intensities, we choose to calculate the overlap of the two Patterson functions. The Fourier coefficients are all real, and, according to convention, will be written $|F_{\mathbf{h}}|^2$, $|F_{\mathbf{p}}|^2$, while $\alpha_{\mathbf{h}}$ and $\alpha_{\mathbf{p}}$ are zero. Translations of the origin are not required; thus $\mathbf{d} = 0$. The symmetry of the Patterson function means that it is always convenient to integrate over a volume U symmetrical about the origin so that $\Omega = 0$. Thus for overlap of the Patterson function without translation, we have

$$R = (U/V^3) \sum_{\mathbf{h}} \sum_{\mathbf{p}} |F_{\mathbf{h}}|^2 |F_{\mathbf{p}}|^2 G_{\mathbf{h}, \mathbf{h}'}. \quad (6)$$

The vector \mathbf{h}' , given by (3), is the position of the reciprocal-lattice point \mathbf{p} after a rotation specified by the matrix $[C]$. For this reason the above expression is referred to as the rotation function.

The result (6) has been given by Patterson (1952) in a context chiefly concerned with *crystallographic* symmetry. He shows that the generalized Faltung integrals, of which R is an example (equation (2)), may be regarded as symmetry functions which show the extent to which given symmetry operations are obeyed by the structure. In this paper, we show that the extension to non-crystallographic symmetry operations has a useful application.

3. The physical significance of the R function

The maximum value of $3(\sin 2\pi x - 2\pi x \cos 2\pi x)/(2\pi x)^3$ is 1.00 (see Fig. 1). It is never greater than 0.086 outside the range $-0.725 < x < +0.725$. Hence all terms in (6) for which $|Hr| > 0.725$ may well be neglected if U is assumed to be a sphere of radius r . Conversely, G will be a maximum when $H = 0$. Now H is the distance of the point $(h_1 + h_1', h_2 + h_2', h_3 + h_3')$ from the origin of reciprocal space. Therefore, from (3), $G = 1$

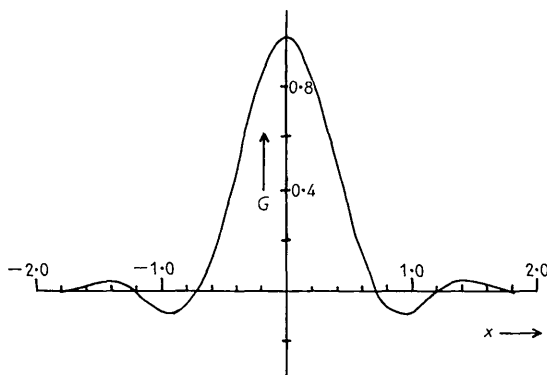


Fig. 1. The interference function

$$G = 3(\sin 2\pi x - 2\pi x \cos 2\pi x)/(2\pi x)^3$$

applicable for roughly spherical sub-units.

for the non-integral values of (p_1, p_2, p_3) which satisfy the three simultaneous equations

$$\begin{aligned} p_1 c_{11} + p_2 c_{12} + p_3 c_{13} &= -h_1 \\ p_1 c_{21} + p_2 c_{22} + p_3 c_{23} &= -h_2 \\ p_1 c_{31} + p_2 c_{32} + p_3 c_{33} &= -h_3. \end{aligned} \quad (7)$$

In other words the integral reciprocal-lattice point h_1, h_2, h_3 has been rotated to the non-integral reciprocal-lattice position (p_1, p_2, p_3) . The summation (6), however, includes terms only for which (p_1, p_2, p_3) are all integral, where $|F_{\mathbf{p}}|^2$ can be measured. At integral values of (p_1, p_2, p_3) close to the non-integral point given by the solution of the above three simultaneous equations, H will be small, and only at such points can G have a large value.

The argument in the preceding sections has been put in terms of the rotation of the self-Patterson function of the sub-units. An identical argument might be given in terms of the rotation of the Fourier transform of the self-Patterson function within an infinite unit cell. This would be a continuous transform. We would rotate the point h_1, h_2, h_3 to the point p_1, p_2, p_3 in order to superimpose the transform of two identical units. The limitations of the finite unit cell require us to interpolate the value of the continuous transform at non-integral points by means of

$$(U/V) \sum_{\mathbf{p}} |F|^2 G_{\mathbf{h}, \mathbf{h}'}$$

The actual number of significant terms in the summation over \mathbf{p} depends on the rapidity with which the transform changes.

If r is large, causing Hr to increase rapidly as H increases, then most of the terms in the summation are negligible. That is, when r is large the rate of change of the transform between reciprocal-lattice points is small, necessitating the inclusion of fewer terms in the interpolation process. Generally r designates the limits of the self-Patterson function, so that the size of the sub-unit determines the number of terms in the interpolation summation.

4. Rotation of a set of oblique axes

In order to calculate the rotation function R corresponding to any desired rotation, all that remains is to calculate the corresponding matrix $[C]$. This problem has been treated from the crystallographic standpoint by Patterson (1959). The crystallographic axes will in general be unequal and oblique, and the least clumsy procedure appears to be:

- (i) to transform the coordinates to a Cartesian form;
- (ii) to transform the resulting coordinates to a rotated set of axes*;

* Note that new coordinates are defined by reference to a set of axes which has been rotated right-handedly. This means that the new coordinates, used with the original axes, result in *left-handed* rotation of the structure. The opposite convention was used in Patterson's (1959) paper, and this has been adjusted in Table 1(b).

Table I

(a) Matrix ρ in terms of Eulerian angles $\theta_1, \theta_2, \theta_3$

$-\sin \theta_1 \cos \theta_2 \sin \theta_3$ $+\cos \theta_1 \cos \theta_3$	$\cos \theta_1 \cos \theta_2 \sin \theta_3$ $+\sin \theta_1 \cos \theta_3$	$\sin \theta_2 \sin \theta_3$
$-\sin \theta_1 \cos \theta_2 \cos \theta_3$ $-\cos \theta_1 \sin \theta_3$	$\cos \theta_1 \cos \theta_2 \cos \theta_3$ $-\sin \theta_1 \sin \theta_3$	$\sin \theta_2 \cos \theta_3$
$\sin \theta_1 \sin \theta_2$	$-\cos \theta_1 \sin \theta_2$	$\cos \theta_2$

(b) Matrix ρ in terms of rotation angle κ and the spherical polar coordinates ψ, φ

$\cos \kappa$ $+\sin^2 \psi \cos^2 \varphi (1 - \cos \kappa)$	$\sin \psi \cos \psi \cos \varphi (1 - \cos \kappa)$ $-\sin \psi \sin \varphi \sin \kappa$	$-\sin^2 \psi \cos \varphi \sin \varphi (1 - \cos \kappa)$ $-\cos \psi \sin \kappa$
$\sin \psi \cos \psi \cos \varphi (1 - \cos \kappa)$ $+\sin \psi \sin \varphi \sin \kappa$	$\cos \kappa$ $+\cos^2 \psi (1 - \cos \kappa)$	$-\sin \psi \cos \psi \sin \varphi (1 - \cos \kappa)$ $+\sin \psi \cos \varphi \sin \kappa$
$-\sin^2 \psi \sin \varphi \cos \varphi (1 - \cos \kappa)$ $+\cos \psi \sin \kappa$	$-\sin \psi \cos \psi \sin \varphi (1 - \cos \kappa)$ $-\sin \psi \cos \varphi \sin \kappa$	$\cos \kappa$ $+\sin^2 \psi \sin^2 \varphi (1 - \cos \kappa)$

(iii) to return these coordinates to their crystallographic form.

In matrix notation, this means

$$[\mathbf{C}] = [\boldsymbol{\alpha}][\boldsymbol{\rho}][\boldsymbol{\beta}]. \quad (8)$$

In this equation

$$[\boldsymbol{\beta}] = \begin{pmatrix} a_1 \sin \alpha_3 \sin \omega & 0 & 0 \\ a_1 \cos \alpha_3 & a_2 & a_3 \cos \alpha_1 \\ a_1 \sin \alpha_3 \cos \omega & 0 & a_3 \sin \alpha_1 \end{pmatrix}, \quad (9)$$

is the matrix which transforms the oblique crystallographic fractional coordinates x_i to Cartesian coordinates X_i . α_i are the crystallographic interaxial angles and $\sin \omega = (\cos \alpha_2 - \cos \alpha_1 \cos \alpha_3) / \sin \alpha_1 \sin \alpha_3$. We have chosen to retain the direction of the X_2 axis along \mathbf{a}_2 , and to set X_1 along $\mathbf{a}_2 \times \mathbf{a}_3$, for convenience in the monoclinic system (see Fig. 2).

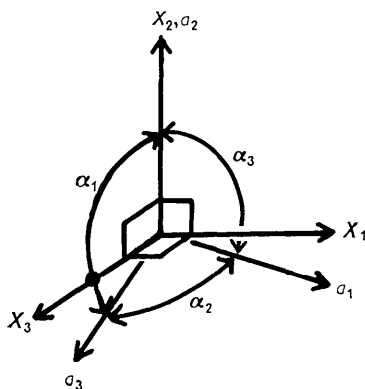


Fig. 2. Relationship of the orthogonal axes X_1, X_2, X_3 to the crystallographic axes x_1, x_2, x_3 .

$$[\boldsymbol{\alpha}] = \begin{pmatrix} 1/a_1 \sin \alpha_3 \sin \omega & 0 & 0 \\ 1/a_2 \tan \alpha_1 \tan \omega - 1/a_2 \tan \alpha_3 \sin \omega & 1/a_2 & -1/a_2 \tan \alpha_1 \\ -1/a_3 \sin \alpha_1 \tan \omega & 0 & 1/a_3 \sin \alpha_1 \end{pmatrix} \quad (10)$$

is the inverse of $[\boldsymbol{\beta}]$. In the case where the rotation function is used to compare the orientation of units

in two different crystal lattices, the a_i and α_i in $[\boldsymbol{\alpha}]$ and $[\boldsymbol{\beta}]$ are different.

The matrix $[\boldsymbol{\rho}]$ (Table I) specifies the rotation operation. $[\boldsymbol{\rho}]$ depends on three variables, and in this work, we have used two different forms of them.

A rotation can be specified by the three Eulerian angles $\theta_1, \theta_2, \theta_3$, whose significance can be most readily seen by reference to Fig. 3. The rotation operation $[\boldsymbol{\rho}]$ consists of (i) a rotation of the Cartesian axes by θ_1 about the X_3 axis, (ii) a rotation θ_2 about the new position of the X_1 axis, and (iii) a rotation θ_3 about the new X_3 axis. These angles are positive if they are clockwise when looking along the relevant axis, as in the usual right-handed convention.* Table I(a) gives $[\boldsymbol{\rho}]$ in terms of Eulerian angles.

The Eulerian angles are somewhat difficult to visualise, and are preferred only because they show up the symmetry of the rotation function in a convenient way, as is shown in the final section. It is equally possible to specify a single rotation by an

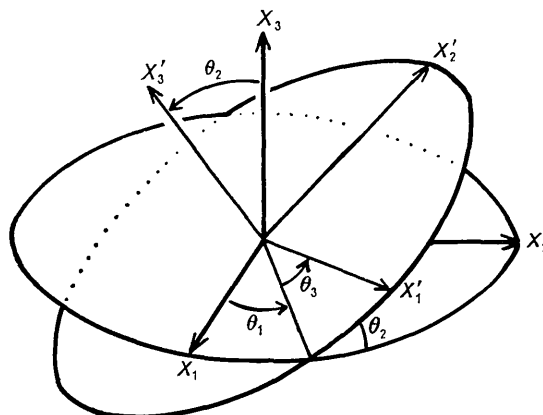


Fig. 3. The Eulerian angles $\theta_1, \theta_2, \theta_3$ relating the rotated axes X'_1, X'_2, X'_3 to the original unrotated orthogonal axes X_1, X_2, X_3 .

* This convention agrees with that of Goldstein (1951), who gives a clear account of the Eulerian angles, together with an interesting discussion of the different sign conventions adopted by various authors.

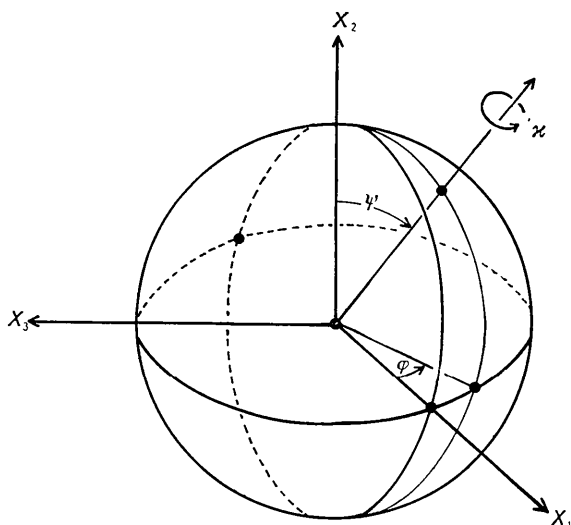


Fig. 4. The variables ψ and φ , polar coordinates which specify a direction about which the axes may be rotated through an angle κ .

angle κ about a given axis. Patterson (1959) has quoted the $[\rho]$ matrix in terms of this rotation and the direction cosines of the rotation axis. In the matrix of Table 1(b) we have reduced this to a system of three variables by writing the direction cosines in terms of spherical polar coordinates ψ , φ of the rotation axis (Fig. 4). For convenience in the monoclinic system, we have retained X_2 as the unique axis, while κ , ψ and φ again follow the right-handed convention.

The relationship between the two sets of variables may be established by comparison of the elements of the two matrices (Table 1(a) and (b)). One finds

$$\begin{aligned} \cos(\kappa/2) &= \cos(\theta_2/2) \cos\left(\frac{\theta_1 + \theta_3}{2}\right) \\ \tan \varphi &= -\cot(\theta_2/2) \sin\left(\frac{\theta_1 + \theta_3}{2}\right) \sec\left(\frac{\theta_1 - \theta_3}{2}\right) \\ \cos \varphi \tan \psi &= \cot\left(\frac{\theta_1 - \theta_3}{2}\right). \end{aligned} \quad (11)$$

Since φ and ψ can always be chosen in the range 0 to π , these equations suffice to find (κ, ψ, φ) from any set $(\theta_1, \theta_2, \theta_3)$.

5. Application the horse haemoglobin

Horse haemoglobin crystallizes in space group $C2$ with half a molecule per asymmetric unit. The two halves of the molecule are related by a crystallographic two-fold axis. Perutz *et al.* (1960) have shown that each half molecule consists of two similar configurations, each representing approximately the structure of myoglobin. Although these are not quite identical, the rotation function showed accurately their relative orientation.

About 1000 independent reflexions with a spacing greater than approximately 6 Å were used. The volume U for integration was a parallelepiped specified by the limits $x_1 = \pm a_1/4$, $x_2 = \pm a_2/2$, $x_3 = \pm a_3/2$ in the unrotated Patterson. For any one set of Miller indices (h_1, h_2, h_3) eight different sets of (p_1, p_2, p_3) were considered for interpolation. These eight reciprocal-lattice points lay at the corners of the reciprocal-lattice unit cell containing the non-integral position given by solution of (7). However, four of these eight points were always on the site of systematic absences. Thus, essentially, a four-point interpolation was used to determine the amplitude of the transform at the non-integral position given by (3).

The value of each intensity was packed in the store of the computer EDSAC 2 as ten binary bits. In this way all 1000 intensities could be accommodated in just over half the free store. The position of any one intensity in the store was entirely determined by its Miller indices. This method of packing not only avoided listing of indices, but avoided time-consuming hunting procedures to find any particular $|F_p|^2$ value. It is important to note that, when (h_1, h_2, h_3) rotates to the non-integral position (h_1', h_2', h_3') , the point $(\bar{h}_1, \bar{h}_2, \bar{h}_3)$ related by the monoclinic symmetry does not rotate to $(\bar{h}_1', \bar{h}_2', \bar{h}_3')$. Thus a single summation involved approximately 2000 independent values of $|F_h|^2$, each multiplied by the appropriate summation

$$\sum_{\mathbf{p}} |F_{\mathbf{p}}|^2 G_{\mathbf{h}, \mathbf{h}'}$$

taken over four terms. The 8000-term summation needed to determine the value of R for a particular set of Eulerian angles took the machine $3\frac{3}{4}$ minutes.

Expression (6) shows that the rotation function, R , is a necessarily positive function, chiefly because of overlap of the origin peaks of the Pattersons. Values of R well away from the important peaks are mostly in the range 380 ± 40 units on the scale adopted. The origin peak of R (no rotation) is 621 units, and the peak at $(85^\circ, 40^\circ, 95^\circ)$ is 504 units. After subtracting the mean background, this peak is about half the origin peak, and considerably greater than the random fluctuations.

The results for points around the largest peak are shown in Fig. 5. Its maximum occurs at $\theta_1 = 85^\circ$, $\theta_2 = 40^\circ$, $\theta_3 = 95^\circ$. The amount of rotation, and the direction of the rotation axes, can be found by substituting these, and symmetry-related Eulerian angles, into expressions (11). The following two pairs of axes are then found:

$$\begin{aligned} A_1, A_2: \kappa &= 180^\circ \pm 3^\circ, \quad \psi = 90^\circ, \quad \varphi = 20^\circ. \\ A_3, A_4: \kappa &= 180^\circ, \quad \psi = 90^\circ \pm 2^\circ, \quad \varphi = 110^\circ. \end{aligned}$$

These four axes, whose positions are plotted on a stereogram in Fig. 6, are inter-related by the crystallographic two-fold axis. They are themselves close to being two-fold axes, and with the crystallographic

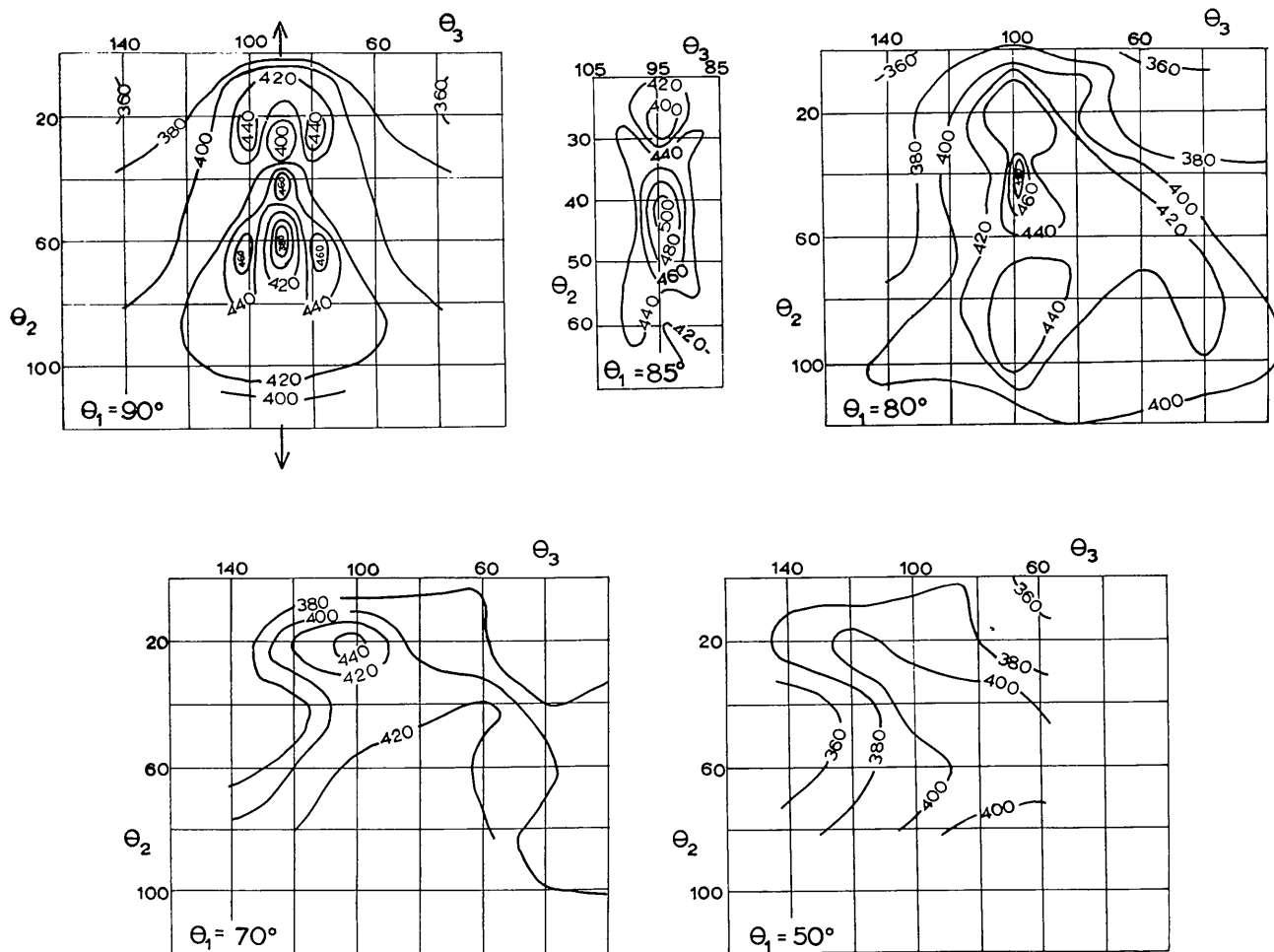


Fig. 5. A plot of the R function, for horse haemoglobin, in terms of the Eulerian angles θ_1 , θ_2 , θ_3 in the vicinity of the large peak. The peak corresponds to the non-crystallographic operation that rotates the α chain into the β chain. Heights are given on an arbitrary scale: one unit corresponds to $26.8 \text{ e.}^4/\text{\AA}^6$. The origin has a height of 621 units.

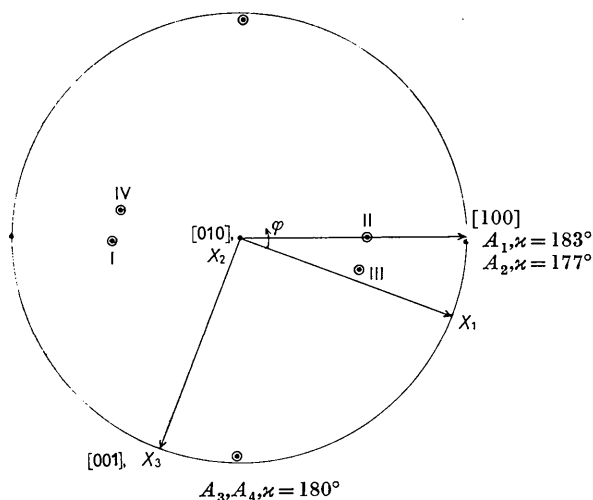


Fig. 6. Stereogram showing the directions of the non-crystallographic rotation axes $A_1 \dots A_4$ of the horse haemoglobin molecule. The haem normal directions, I, II, III and IV, found from electron spin resonance data, are also shown. Points above the XZ plane are indicated by dots, points below by circles.

axis, form an almost orthogonal set. If the axes all passed through one point, the molecules would very nearly have the point group 222.

There are two sets of measurements with which these results may be compared. Associated with each quarter of the haemoglobin molecule is a planar haem group, whose orientation has been determined by electron-spin resonance (Ingram *et al.*, 1956). The directions of the haem normals are shown in Fig. 6, and are consistent within 6° with the directions of the pseudo axes found by the rotation function. Cullis *et al.* (1961) attempted to determine the direction of the pseudo rotation axes by examination of the polypeptide chain directions in the 6 \AA resolution Fourier map of haemoglobin. The attempt showed up clearly the slight differences between the quarter-molecules, but the result agrees within 6° with the position of the rotation function peak.

The rotation matrix $[C]$ corresponding to the peak of R is

$$\begin{pmatrix} -1.0098 & -0.0126 & +0.3429 \\ +0.0014 & -0.9982 & -0.0485 \\ -0.0573 & -0.0736 & +1.0080 \end{pmatrix}$$

6. Resolution

The full range of Eulerian angles was not explored, using all the data. It would have taken about 50 hr. on EDSAC 2 to cover all independent Eulerian angles, taken at 20° intervals. Instead, a survey was made using a very restricted amount of data. Each summation was therefore much faster, and a smaller number of points needed to be evaluated. Complete coverage of all reflexions with spacing over 20 \AA and at 30° intervals of the Eulerian angles took under 1 hr. The highest non-origin peak fell in the same region as the maximum of the rotation function, found by using all 1000 independent reflexions.

It is useful to know the smallest change of Eulerian angle which will give a significant effect. Roughly speaking, it is that rotation which moves the most distant reciprocal-lattice points through one reciprocal-lattice translation. If h is the highest h_1 , h_2 or h_3 index, angles less than $2 \sin^{-1}(1/2h)$ are insignificant. This gives, in our case, 3.5° ; the closest angular intervals used, namely 5° , are of the same order. A survey around the peak was first done with larger intervals of 20° and 10° ; 5° intervals were used only to obtain the exact peak position.

The peak may be 'sharpened' in exactly the same way as a normal Fourier synthesis, by weighting down, or omitting, the inner reflexions of the reciprocal lattice. As an experiment, only those reflexions between 10 \AA and 6 \AA spacing were used. Fig. 7 shows a line section passing near the peak using this partial data, and, for comparison, all the data. The partial data shows greater background variations because of the smaller number of reflexions involved.

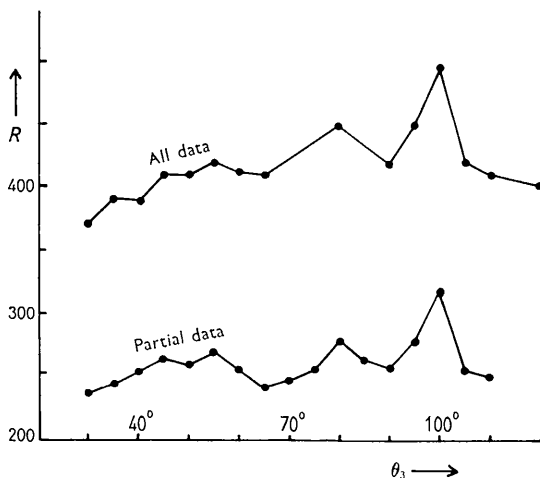


Fig. 7. A line section of $R(\theta_1=80^\circ, \theta_2=40^\circ, \theta_3)$ passing near the large peak. Upper line: All reflexions, with spacing greater than 6 \AA . Lower line: Only reflexions with spacing between 10 and 6 \AA included.

7. Symmetry properties of the rotation function, R

As the amount of computing required is critical, it is particularly important to know what range of angles

needs to be explored before all independent rotation operations have been considered. This will depend on the point group of the rotated object, and in this final section, we explain how this range may be determined.

Although there are compelling reasons (which will appear) for carrying out the computation in terms of the Eulerian angles $(\theta_1, \theta_2, \theta_3)$, it is often much easier to think of the rotation in terms of (κ, ψ, φ) . If the $(\theta_1, \theta_2, \theta_3)$ cannot be visualized directly, they can be interpreted in terms of (κ, ψ, φ) by the use of equations (11), or by comparing the elements of the matrix $[\rho]$.

In both systems, a given rotation operation can have several different expressions:

(a) If any angular variable lies outside the range 0 to 2π , it can be brought into this range by adding an integral multiple of 2π , without affecting the rotation operation. Thus we can write

$$[\rho(\kappa, \psi, \varphi)] \equiv [\rho(\kappa + 2\pi n_1, \psi + 2\pi n_2, \varphi + 2\pi n_3)],$$

where n_1, n_2, n_3 are integers.

(b) Referring to Fig. 4, if ψ is greater than π the rotation operation is the same as one in which φ is increased by π and ψ becomes $2\pi - \psi$. Thus

$$[\rho(\kappa, \psi, \varphi)] \equiv [\rho(\kappa, 2\pi - \psi, \varphi + \pi)].$$

(c) Note that a rotation κ about any axis is equivalent to a rotation $-\kappa$ about an oppositely directed axis. Thus

$$\rho(\kappa, \psi, \varphi) \equiv \rho(-\kappa, \pi - \psi, \pi + \varphi).$$

All these identities can be checked by substituting into $[\rho]$ (Table 1(b)), and seeing that no element is changed.

All rotation operations are therefore included in

$$\begin{aligned} 0 &\leq \kappa < 2\pi, \\ 0 &\leq \psi \leq \pi, \\ 0 &\leq \varphi < \pi. \end{aligned}$$

The corresponding relationships in $(\theta_1, \theta_2, \theta_3)$ are:

$$\begin{aligned} [\rho(\theta_1, \theta_2, \theta_3)] &\equiv [\rho(\theta_1 + 2\pi n_1, \theta_2 + 2\pi n_2, \theta_3 + 2\pi n_3)], \\ [\rho(\theta_1, \theta_2, \theta_3)] &\equiv [\rho(\theta_1 + \pi, -\theta_2, \theta_3 + \pi)]. \end{aligned}$$

The full range of rotation operations is

$$\begin{aligned} 0 &\leq \theta_1 < \pi \\ 0 &\leq \theta_2 < 2\pi \\ 0 &\leq \theta_3 < 2\pi. \end{aligned}$$

So far, we have considered physically identical operations, which lead to identical expressions of the $[\rho]$ matrix. In addition, point-group symmetry in the rotated object (in our case, the reciprocal lattice) will cause the same value of the rotation function to be found for physically distinct rotations. The point group of the haemoglobin reciprocal lattice ($2/m$) will be used for illustration.

(d) Equal rotations about directions related by the point group symmetry will give the same result.

Thus, in haemoglobin, where the dyad relates points with coordinates φ and $\varphi + \pi$, we have

$$R(\kappa, \psi, \varphi) = R(\kappa, \psi, \varphi + \pi).$$

Taking this in conjunction with (c), we find that the mirror-plane symmetry

$$R(\kappa, \psi, \varphi) = R(-\kappa, \pi - \psi, \varphi)$$

is automatically produced, showing that the rotation operations have the same symmetry for point groups 2 and $2/m$. This is true because only *proper* rotations (rotation without inversion) are considered.

The corresponding relationship for the Eulerian angles is

$$R(\theta_1, \theta_2, \theta_3) = R(\pi - \theta_1, \theta_2, \pi - \theta_3).$$

(e) Two consecutive rotations about different axes are always equivalent to a single rotation about some other axis, given by multiplication of the two rotation matrices. If $[\rho_1]$ represents the crystallographic rotation of π about \mathbf{a}_2 , and $[\rho]$ a general rotation (κ, ψ, φ) , then the rotation $[\rho][\rho_1]$ is a complicated function of κ, ψ, φ . However, in Eulerian angles such relationships take a simple form. For a two-fold rotation about X_2 before a general rotation one finds

$$R(\theta_1, \theta_2, \theta_3) = R(\theta_1, \theta_2 + \pi, \pi - \theta_3),$$

so that the range of rotations can be expressed quite simply. For this reason the required range of angles can only be expressed conveniently in terms of the Eulerian angles, and for this reason they are preferred for computation.

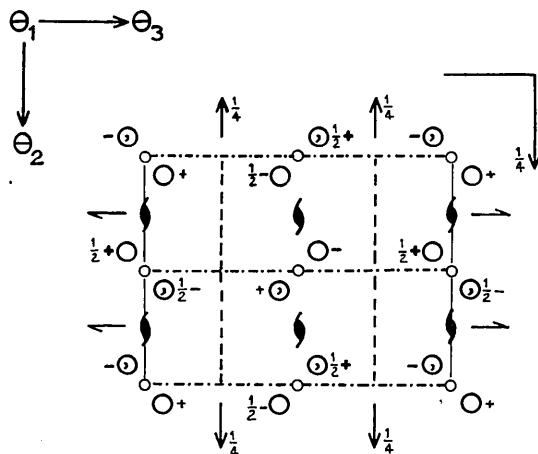


Fig. 8. Diagram showing equivalent values of the Eulerian angles for rotations of an object with point group symmetry $2/m$.

To collect all these relationships together in a form familiar to crystallographers, it is convenient to consider the symmetry of R in a three-dimensional space with orthogonal coordinates $\theta_1, \theta_2, \theta_3$. Postulate (a) above tells us that equivalent points lie on a lattice with spacing 2π in each direction, and the other relationships may be indicated by constructing a conventional space-group diagram (Fig. 8). The space group in this case is No. 56 ($Pbn\bar{b}$ retaining the order $\theta_1, \theta_2, \theta_3$). The size of the asymmetric unit, giving the range of angles which needs to be explored, may be expressed

$$\begin{aligned} 0 &\leq \theta_1 \leq \pi/2 \\ 0 &\leq \theta_2 < \pi \\ 0 &\leq \theta_3 < 2\pi. \end{aligned}$$

We are indebted to Dr M. F. Perutz for allowing us to use the three-dimensional 6 Å haemoglobin data, and to the Director of the University of Cambridge Mathematical Laboratory for making the EDSAC 2 computer available. We are grateful for most useful discussions with Dr V. Heine and Dr A. Klug, and to Mrs D. Thomas who has given assistance in the preparation of this manuscript.

References

- BODO, G., DINTZIS, H. M., KENDREW, J. C. & WYCKOFF, H. W., (1959). *Proc. Roy. Soc. A* **253**, 70.
 CASPAR, D. L. D. (1956). *Nature, Lond.* **177**, 475.
 CULLIS, A. F., MUIRHEAD, H., NORTH, A. T., PERUTZ, M. F. & ROSSMANN, M. G. (1961). *Proc. Roy. Soc.* (In press.)
 CRICK, F. H. C. & WATSON, J. D. (1956). *Nature, Lond.* **177**, 473.
 FINCH, J. T. & KLUG, A. (1959). *Nature, Lond.* **183**, 1709.
 FRANKLIN, R. E. & HOLMES, K. C. (1958). *Acta Cryst.* **11**, 213.
 GOLDSTEIN, H. (1951). *Classical Mechanics*, p. 107. Cambridge, Mass.: Addison-Wesley Press.
 HARRISON, P. M. (1959). *J. Mol. Biol.* **1**, 69.
 INGRAM, D. J. E., GIBSON, J. F. & PERUTZ, M. F. (1956). *Nature, Lond.* **178**, 907.
 KLUG, A. & FINCH, J. T. (1960). *J. Mol. Biol.* **2**, 201.
 MAGDOFF, B. S. (1960). *Nature, Lond.* **185**, 673.
 PATTERSON, A. L. (1939). *Phys. Rev.* **56**, 972.
 PATTERSON, A. L. (1952). In *Computing methods and the phase problem*, p. 29. Pennsylvania State College.
 PATTERSON, A. L. (1959). *International Tables for X-ray Crystallography*, **2**, 52. Birmingham: Kynoch Press.
 PERUTZ, M. F., ROSSMANN, M. G., CULLIS, A. F., MUIRHEAD, H., WILL, G. & NORTH, A. C. T. (1960). *Nature, Lond.* **185**, 416.
 PERUTZ, M. F., TROTTER, I. F., HOWELLS, E. R. & GREEN, D. W. (1955). *Acta Cryst.* **8**, 241.
 WATSON, J. D. (1954). *Biochim. Biophys. Acta.* **13**, 10