International Tables for Crystallography (2010). Vol. B, Chapter 2.3, pp. 244–281.

2.3. Patterson and molecular replacement techniques, and the use of noncrystallographic symmetry in phasing

By L. Tong, M. G. Rossmann and E. Arnold

2.3.1. Introduction

2.3.1.1. Background

Historically, the Patterson has been used in a variety of ways to effect the solutions of crystal structures. While some simple structures (Ketelaar & de Vries, 1939; Hughes, 1940; Speakman, 1949; Shoemaker *et al.*, 1950) were solved by direct analysis of Patterson syntheses, alternative methods have largely superseded this procedure. An early innovation was the heavy-atom method which depends on the location of a small number of relatively strong scatterers (Harker, 1936). Image-seeking methods and Patterson superposition techniques were first contemplated in the late 1930s (Wrinch, 1939) and applied sometime later (Beevers & Robertson, 1950; Clastre & Gay, 1950; Garrido, 1950*a*; Buerger, 1959). This experience provided the encouragement for computerized vector-search methods to locate individual atoms automatically (Mighell & Jacobson, 1963; Kraut, 1961; Hamilton, 1965; Simpson *et al.*, 1965) or to position known molecular fragments in unknown crystal structures (Nordman & Nakatsu, 1963; Huber, 1965). The Patterson function has been used extensively in conjunction with the isomorphous replacement method (Rossmann, 1960; Blow, 1958) or anomalous dispersion (Rossmann, 1961*a*) to determine the position of heavy-atom substitution. Pattersons have been used to detect the presence and relative orientation of multiple copies of a given chemical motif in the crystallographic asymmetric unit in the same or different crystals (Rossmann & Blow, 1962). Finally, the orientation and placement of known molecular structures ('molecular replacement') into unknown crystal structures can be accomplished *via* Patterson techniques.

The function, introduced by Patterson in 1934 (Patterson, 1934*a*,*b*), is a convolution of electron density with itself and may be defined as

$$P(\mathbf{u}) = \int_V \rho(\mathbf{x}) \cdot \rho(\mathbf{u}+\mathbf{x})\, \mathrm{d}\mathbf{x}, \qquad (2.3.1.1)$$

where $P(\mathbf{u})$ is the 'Patterson' function at $\mathbf{u}$, $\rho(\mathbf{x})$ is the crystal's periodic electron density and V is the volume of the unit cell. The Patterson function, or F^2 series, can be calculated directly from the experimentally derived X-ray intensities as

$$P(\mathbf{u}) = \frac{2}{V^2} \sum_{\mathbf{h}}^{\text{hemisphere}} |\mathbf{F}_\mathbf{h}|^2 \cos 2\pi\mathbf{h}\cdot\mathbf{u}. \qquad (2.3.1.2)$$

The derivation of (2.3.1.2) from (2.3.1.1) can be found in this volume (see Section 1.3.4.2.1.6) along with a discussion of the physical significance and symmetry of the Patterson function, although the principal properties will be restated here.

The Patterson can be considered to be a vector map of all the pairwise interactions between the atoms in a unit cell. The vectors in a Patterson correspond to vectors in the real (direct) crystal cell but translated to the Patterson origin. Their weights are proportional to the product of densities at the tips of the vectors in the real cell. The Patterson unit cell has the same size as the real crystal cell. The symmetry of the Patterson comprises the Laue point group of the crystal cell plus any additional lattice symmetry due to Bravais centring. The reduction of the real space group to the Laue symmetry is produced by the translation of all vectors to the Patterson origin and the introduction of a centre of symmetry. The latter is a consequence of the relationship between the vectors **AB** and **BA**. The Patterson symmetries for all 230 space groups are tabulated in *IT* A (2005).

An analysis of Patterson peaks can be obtained by considering N atoms with form factors f_i in the unit cell. Then

$$\mathbf{F}_\mathbf{h} = \sum_{i=1}^{N} f_i \exp(2\pi i \mathbf{h}\cdot\mathbf{x}_i).$$

Using Friedel's law,

$$\begin{aligned} |\mathbf{F}_\mathbf{h}|^2 &= \mathbf{F}_\mathbf{h}\cdot\mathbf{F}_\mathbf{h}^* \\ &= \left[\sum_{i=1}^{N} f_i \exp(2\pi i\mathbf{h}\cdot\mathbf{x}_i)\right]\left[\sum_{j=1}^{N} f_j \exp(-2\pi i\mathbf{h}\cdot\mathbf{x}_j)\right], \end{aligned}$$

which can be decomposed to

$$|\mathbf{F}_\mathbf{h}|^2 = \sum_{i=1}^{N} f_i^2 + \sum_{i\neq j}^{N}\sum_{j}^{N} f_i f_j \exp[2\pi i\mathbf{h}\cdot(\mathbf{x}_i-\mathbf{x}_j)]. \qquad (2.3.1.3)$$

On substituting (2.3.1.3) in (2.3.1.2), we see that the Patterson consists of the sum of N^2 total interactions of which N are of weight f_i^2 at the origin and $N(N-1)$ are of weight $f_i f_j$ at $\mathbf{x}_i - \mathbf{x}_j$.

The weight of a peak in a real cell is given by

$$w_i = \int_U \rho_i(\mathbf{x})\, \mathrm{d}\mathbf{x} = Z_i \text{ (the atomic number)},$$

where U is the volume of the atom i. By analogy, the weight of a peak in a Patterson (form factor $f_i f_j$) will be given by

$$w_{ij} = \int_U P_{ij}(\mathbf{u})\, \mathrm{d}\mathbf{u} = Z_i Z_j.$$

Although the maximum height of a peak will depend on the spread of the peak, it is reasonable to assume that heights of peaks in a Patterson are proportional to the products of the atomic numbers of the interacting atoms.

There are a total of N^2 interactions in a Patterson due to N atoms in the crystal cell. These can be represented as an $N\times N$ square matrix whose elements $\mathbf{u}_{ij}$, w_{ij} indicate the position and weight of the peak produced between atoms i and j (Table 2.3.1.1). The N vectors corresponding to the diagonal of this matrix are located at the Patterson origin and arise from the convolution of each atom with itself. This leaves $N(N-1)$ vectors whose locations depend on the relative positions of all of the atoms in the crystal cell and whose weights depend on the atom types related by the vector. Complete specification of the unique non-origin Patterson vectors requires description of only the $N(N-1)/2$ elements in either the upper or the lower triangle of this matrix, since the two sets of vectors represented by the two triangles are related by a centre of symmetry

Table 2.3.1.1. *Matrix representation of Patterson peaks*

The $N \times N$ matrix represents the position $\mathbf{u}_{ij}$ and weights w_{ij} of atomic interactions in a Patterson arising from N atoms at $\mathbf{x}_i$ and weight w_i in the real cell.

	$\mathbf{x}_1, w_1$	$\mathbf{x}_2, w_2$	...	$\mathbf{x}_N, w_N$
$\mathbf{x}_1, w_1$	$\mathbf{u}_{11} = \mathbf{x}_1 - \mathbf{x}_1$, $w_{11} = w_1^2$	$\mathbf{u}_{12} = \mathbf{x}_1 - \mathbf{x}_2$, $w_{12} = w_1 w_2$	...	$\mathbf{u}_{1N} = \mathbf{x}_1 - \mathbf{x}_N$, $w_{1N} = w_1 w_N$
$\mathbf{x}_2, w_2$	$\mathbf{x}_2 - \mathbf{x}_1, w_2 w_1$	$0, w_2^2$	...	$\mathbf{x}_2 - \mathbf{x}_N, w_2 w_N$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
$\mathbf{x}_N, w_N$	$\mathbf{x}_N - \mathbf{x}_1, w_N w_1$	$\mathbf{x}_N - \mathbf{x}_2, w_N w_2$	...	$0, w_N^2$

$[\mathbf{u}_{ij} \equiv \mathbf{x}_i - \mathbf{x}_j = -\mathbf{u}_{ij} \equiv -(\mathbf{x}_j - \mathbf{x}_i)]$. Patterson vector positions are usually represented as $\langle uvw \rangle$, where u, v and w are expressed as fractions of the Patterson cell axes.

2.3.1.2. Limits to the number of resolved vectors

If we assume a constant number of atoms per unit volume, the number of atoms N in a unit cell increases in direct proportion with the volume of the unit cell. Since the number of non-origin peaks in the Patterson function is $N(N-1)$ and the Patterson cell is the same size as the real cell, the problem of overlapping peaks in the Patterson function becomes severe as N increases. To make matters worse, the breadth of a Patterson peak is roughly equal to the sum of the breadth of the original atoms. The effective width of a Patterson peak will also increase with increasing thermal motion, although this effect can be artificially reduced by sharpening techniques. Naturally, a loss of attainable resolution at high scattering angles will affect the resolution of atomic peaks in the real cell as well as peaks in the Patterson cell. If U is the van der Waals volume per average atom, then the fraction of the cell occupied by atoms will be $f = NU/V$. Similarly, the fraction of the cell occupied by Patterson peaks will be $2UN(N-1)/V$ or $2f(N-1)$. With the reasonable assumption that $f \simeq 0.1$ for a typical organic crystal, then the cell can contain at most five atoms ($N \leq 5$) for there to be no overlap, other than by coincidence, of the peaks in the Patterson. As N increases there will occur a background of peaks on which are superimposed features related to systematic properties of the structure.

The contrast of selected Patterson peaks relative to the general background level can be enhanced by a variety of techniques. For instance, the presence of heavy atoms not only enhances the size of a relatively small number of peaks but ordinarily ensures a larger separation of the peaks due to the light-atom skeleton on which the heavy atoms are hung. That is, the factor f (above) is substantially reduced. Another example is the effect of systematic atomic arrangements (*e.g.* α-helices or aromatic rings) resulting in multiple peaks which stand out above the background. In the isomorphous replacement method, isomorphous difference Pattersons are utilized in which the contrast of the Patterson interactions between the heavy atoms is enhanced by removal of the predominant interactions which involve the rest of the structure.

2.3.1.3. Modifications: origin removal, sharpening etc.

A. L. Patterson, in his first in-depth exposition of his newly discovered F^2 series (Patterson, 1935), introduced the major modifications to the Patterson which are still in use today. He illustrated, with one-dimensional Fourier series, the techniques of removing the Patterson origin peak, sharpening the overall function and also removing peaks due to atoms in special positions. Each one of these modifications can improve the interpretability of Pattersons, especially those of simple structures. Whereas the recommended extent of such modifications is controversial (Buerger, 1966), most studies which utilize Patterson functions do incorporate some of these techniques [see, for example, Jacobson *et al.* (1961), Braun *et al.* (1969) and Nordman (1980*a*)]. Since Patterson's original work, other workers have suggested that the Patterson function itself might be modified; Fourier inversion of the modified Patterson then provides a new and perhaps more tractable set of structure factors (McLachlan & Harker, 1951; Simonov, 1965; Raman, 1966; Corfield & Rosenstein, 1966). Karle & Hauptman (1964) suggested that an improved set of structure factors could be obtained from an origin-removed Patterson modified such that it was everywhere non-negative and that Patterson density values less than a bonding distance from the origin were set to zero. Nixon (1978) was successful in solving a structure which had previously resisted solution by using a set of structure factors which had been obtained from a Patterson in which the largest peaks had been attenuated.

The N origin peaks [see expression (2.3.1.3)] may be removed from the Patterson by using coefficients

$$|\mathbf{F}_{\mathbf{h},\,\mathrm{mod}}|^2 = |\mathbf{F}_{\mathbf{h}}|^2 - \sum_{i=1}^{N} f_i^2.$$

A Patterson function using these modified coefficients will retain all interatomic vectors. However, the observed structure factors $\mathbf{F}_{\mathbf{h}}$ must first be placed on an absolute scale (Wilson, 1942) in order to match the scattering-factor term. In practice, Patterson origins can also be removed by using coefficients

$$\left|\mathbf{F}_{\mathbf{h},\mathrm{mod}}\right|^2 = \left|\mathbf{F}_{\mathbf{h}}\right|^2 - \langle \left|\mathbf{F}_{\mathbf{h}}\right|^2 \rangle,$$

where $\langle \left|\mathbf{F}_{\mathbf{h}}\right|^2 \rangle$ is the average reflection intensity, usually calculated in several resolution shells. This formula has the advantage that the observed structure factors do not need to be on absolute scale.

Analogous to origin removal, the vector interactions due to atoms in known positions can also be removed from the Patterson function. Patterson showed that non-origin Patterson peaks arising from known atoms 1 and 2 may be removed by using the expression

$$|\mathbf{F}_{\mathbf{h},\,\mathrm{mod}}|^2 = |\mathbf{F}_{\mathbf{h}}|^2 - \sum_{i=1}^{N} f_i^2 t_i^2 - 2 f_1 f_2 t_1 t_2 \cos 2\pi \mathbf{h} \cdot (\mathbf{x}_1 - \mathbf{x}_2),$$

where $\mathbf{x}_1$ and $\mathbf{x}_2$ are the positions of atoms 1 and 2 and t_1 and t_2 are their respective thermal correction factors. Using one-dimensional Fourier series, Patterson illustrated how interactions due to known atoms can obscure other information.

Patterson also introduced a means by which the peaks in a Patterson function may be artificially sharpened. He considered the effect of thermal motion on the broadening of electron-density peaks and consequently their Patterson peaks. He suggested that the F^2 coefficients could be corrected for thermal effects by simulating the atoms as point scatterers and proposed using a modified set of coefficients

$$|\mathbf{F}_{\mathbf{h},\,\mathrm{sharp}}|^2 = |\mathbf{F}_{\mathbf{h}}|^2 / \bar{f}^2,$$

where $\bar{f}$, the average scattering factor per electron, is given by

$$\bar{f} = \sum_{i=1}^{N} f_i \Big/ \sum_{i=1}^{N} Z_i.$$

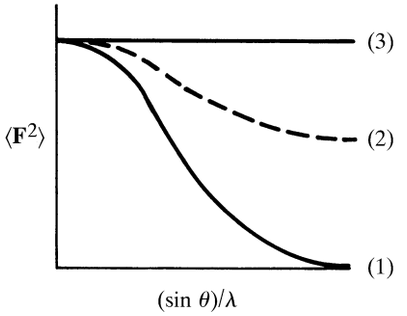


Fig. 2.3.1.1. Effect of 'sharpening' Patterson coefficients. (1) shows a mean distribution of $|\mathbf{F}|^2$ values with resolution, $(\sin\theta)/\lambda$. The normal decline of this curve is due to increasing destructive interference between different portions within diffuse atoms at larger Bragg angles. (2) shows the distribution of 'sharpened' coefficients. (3) shows the theoretical distribution of $|\mathbf{F}|^2$ produced by a point-atom structure. To represent such a structure with a Fourier series would require an infinite series in order to avoid large errors caused by truncation.

A common formulation for this type of sharpening expresses the atomic scattering factors at a given angle in terms of an overall isotropic thermal parameter B as

$$f(s) = f_0 \exp(-Bs^2).$$

The Patterson coefficients then become

$$\mathbf{F}_{\mathbf{h},\,\mathrm{sharp}} = \frac{Z_{\mathrm{total}}}{\sum_{i=l}^{N} f(s)} \mathbf{F}_{\mathbf{h}}.$$

The normalized structure factors, E (see Chapter 2.2), which are used in crystallographic direct methods, are also a common source of sharpened Patterson coefficients $(E^2 - 1)$. Although the centre positions and total contents of Patterson peaks are unaltered by sharpening, the resolution of individual peaks may be enhanced. The degree of sharpening can be controlled by adjusting the size of the assumed B factor; Lipson & Cochran (1966, pp. 165–170) analysed the effect of such a choice on Patterson peak shape.

All methods of sharpening Patterson coefficients aim at producing a point atomic representation of the unit cell. In this quest, the high-resolution terms are enhanced (Fig. 2.3.1.1). Unfortunately, this procedure must also produce a serious Fourier truncation error which will be seen as large ripples about each Patterson peak (Gibbs, 1898). Consequently, various techniques have been used (mostly unsuccessfully) in an attempt to balance the advantages of sharpening with the disadvantages of truncation errors.

Schomaker and Shoemaker [unpublished; see Lipson & Cochran (1966, p. 168)] used a function

$$|\mathbf{F}_{\mathbf{h},\,\mathrm{sharp}}|^2 = \frac{|\mathbf{F}_{\mathbf{h}}|^2}{\bar{f}^2} s^2 \exp\left[-\frac{\pi^2}{p} s^2\right],$$

in which s is the length of the scattering vector, to produce a Patterson synthesis which is less sensitive to errors in the low-order terms. Jacobson *et al.* (1961) used a similar function,

$$|\mathbf{F}_{\mathbf{h},\,\mathrm{sharp}}|^2 = \frac{|\mathbf{F}_{\mathbf{h}}|^2}{\bar{f}^2} (k + s^2) \exp\left[-\frac{\pi}{p} s^2\right],$$

which they rationalize as the sum of a scaled exponentially sharpened Patterson and a gradient Patterson function (the value of k was empirically chosen as $\frac{2}{3}$). This approach was subsequently further developed and generalized by Wunderlich (1965).

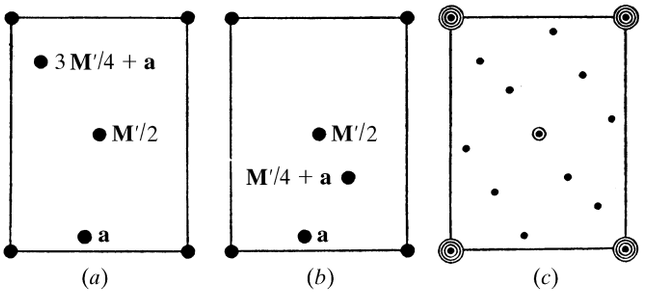


Fig. 2.3.1.2. (*c*) The point Patterson of the two homometric structures in (*a*) and (*b*). The latter are constructed by taking points at $\mathbf{a}$ and $\frac{1}{2}\mathbf{M}'$, where $\mathbf{M}'$ is a cell diagonal, and adding a third point which is (*a*) at $\frac{3}{4}\mathbf{M}' + \mathbf{a}$ or (*b*) at $\frac{1}{4}\mathbf{M}' + \mathbf{a}$. [Reprinted with permission from Patterson (1944).]

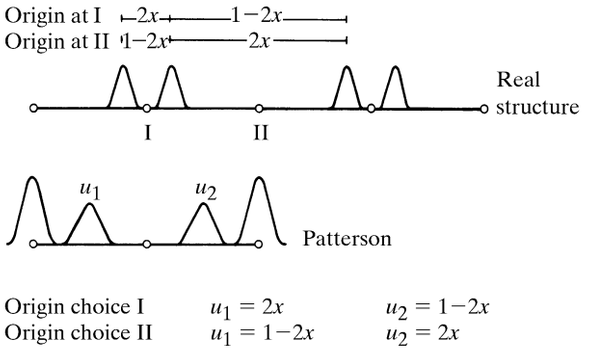
2.3.1.4. Homometric structures and the uniqueness of structure solutions; enantiomorphic solutions

Interpretation of any Patterson requires some assumption, such as the existence of discrete atoms. A complete interpretation might also require an assumption of the number of atoms and may require other external information (*e.g.* bond lengths, bond angles, van der Waals separations, hydrogen bonding, positive density *etc.*). To what extent is the solution of a Patterson function unique? Clearly the greater is the supply of external information, the fewer will be the number of possible solutions. Other constraints on the significance of a Patterson include the error involved in measuring the observed coefficients and the resolution limit to which they have been observed. The larger the error, the larger the number of solutions. When the error on the amplitudes is infinite, it is only the other physical constraints, such as packing, which limit the structural solutions. Alternative solutions of the same Patterson are known as homometric structures.

During their investigation of the mineral bixbyite, Pauling & Shappell (1930) made the disturbing observation that there were two solutions to the structure, with different arrangements of atoms, which yielded the same set of calculated intensities. Specifically, atoms occupying equipoint set 24*d* in space group $I(2_1/a)\bar{3}$ can be placed at either $x, 0, \frac{1}{4}$ or $-x, 0, \frac{1}{4}$ without changing the calculated intensities. Yet the two structures were not chemically equivalent. These authors resolved the ambiguity by placing the oxygen atoms in question at positions which gave the most acceptable bonding distances with the rest of the structure.

Patterson interpreted the above ambiguity in terms of the F^2 series: the distance vector sets or Patterson functions of the two structures were the same since each yielded the same calculated intensities (Patterson, 1939). He defined such a pair of structures a homometric pair and called the degenerate vector set which they produced a homometric set. Patterson went on to investigate the likelihood of occurrence of homometric structures and, indeed, devoted a great deal of his time to this matter. He also developed algebraic formalisms for examining the occurrence of homometric pairs and multiplets in selected one-dimensional sets of points, such as cyclotomic sets, and also sets of points along a line (Patterson, 1944). Some simple homometric pairs are illustrated in Fig. 2.3.1.2.

Drawing heavily from Patterson's inquiries into the structural uniqueness allowed by the diffraction data, Hosemann, Bagchi and others have given formal definitions of the different types of homometric structures (Hosemann & Bagchi, 1954). They suggested a classification divided into pseudohomometric structures and homomorphs, and used an integral equation representing a convolution operation to express their examples of finite homometric structures. Other workers have chosen various



Origin choice I	$u_1 = 2x$	$u_2 = 1-2x$
Origin choice II	$u_1 = 1-2x$	$u_2 = 2x$

Fig. 2.3.2.1. Origin selection in the interpretation of a Patterson of a one-dimensional centrosymmetric structure.

means for describing homometric structures [Buerger (1959, pp. 41–50), Menzer (1949), Bullough (1961, 1964), Hoppe (1962)].

Since a Patterson function is centrosymmetric, the Pattersons of a crystal structure and of its mirror image are identical. Thus the enantiomeric ambiguity present in noncentrosymmetric crystal structures cannot be overcome by using the Patterson alone and represents a special case of homometric structures. Assignment of the correct enantiomorph in a crystal structure analysis is generally not possible unless a recognizable fragment of known chirality emerges (*e.g.* L-amino acids in proteins, D-riboses in nucleic acids, the known framework of steroids and other natural products, the right-handed twist of α-helices, the left-handed twist of successive strands in a β-sheet, the fold of a known protein subunit *etc.*) or anomalous-scattering information is available and can be used to resolve the ambiguity (Bijvoet *et al.*, 1951).

It is frequently necessary to select arbitrarily one enantiomorph over another in the early stages of a structure solution. Structure-factor phases calculated from a single heavy atom in space group $P1$, $P2$ or $P2_1$ (for instance) will be centrosymmetric and both enantiomorphs will be present in Fourier calculations based on these phases. In other space groups (*e.g.* $P2_12_12_1$), the selected heavy atom is likely to be near one of the planes containing the 2_1 axes and thus produce a weaker 'ghost' image of the opposite enantiomorph. The mixture of the two overlapped enantiomorphic solutions can cause interpretive difficulties. As the structure solution progresses, the 'ghosts' are exorcized owing to the dominance of the chosen enantiomorph in the phases.

2.3.1.5. The Patterson synthesis of the second kind

Patterson also defined a second, less well known, function (Patterson, 1949) as

$$P_{\pm}(\mathbf{u}) = \int \rho(\mathbf{u}+\mathbf{x}) \cdot \rho(\mathbf{u}-\mathbf{x})\ \mathrm{d}\mathbf{x}$$
$$= \frac{2}{V^2} \sum_{\mathbf{h}}^{\text{hemisphere}} F_{\mathbf{h}}^2 \cos(2\pi 2\mathbf{h}\cdot\mathbf{u} - 2\alpha_{\mathbf{h}}).$$

This function can be computed directly only for centrosymmetric structures. It can be calculated for noncentrosymmetric structures when the phase angles are known or assumed. It will represent the degree to which the known or assumed structure has a centre of symmetry at $\mathbf{u}$. That is, the product of the density at $\mathbf{u}+\mathbf{x}$ and $\mathbf{u}-\mathbf{x}$ is large when integrated over all values $\mathbf{x}$ within the unit cell. Since atoms themselves have a centre of symmetry, the function will contain peaks at each atomic site roughly proportional in height to the square of the number of electrons in each atom plus peaks at the midpoint between atoms proportional in height to the product of the electrons in each atom. Although this function has not been found very useful in practice, it is useful for demonstrating the presence of weak enantiomorphic images in a given tentative structure determination.

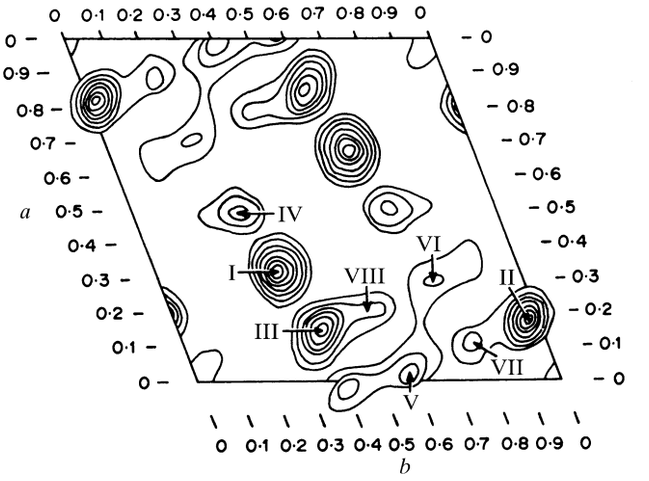


Fig. 2.3.2.2. The *c*-axis projection of cuprous chloride azomethane complex ($C_2H_6Cl_2Cu_2N_2$). The space group is $P\bar{1}$ with one molecule per unit cell. [Adapted from and reprinted with permission from Woolfson (1970, p. 321).]

2.3.2. Interpretation of Patterson maps

2.3.2.1. Simple solutions in the triclinic cell. Selection of the origin

A hypothetical one-dimensional centrosymmetric crystal structure containing an atom at x and at $-x$ and the corresponding Patterson is illustrated in Fig. 2.3.2.1. There are two different centres of symmetry which may be chosen as convenient origins. If the atoms are of equal weight, we expect Patterson vectors at positions $u = \pm 2x$ with weights equal to half the origin peak. There are two symmetry-related peaks, u_1 and u_2 (Fig. 2.3.2.1) in the Patterson. It is an arbitrary choice whether $u_1 = 2x$ or $u_2 = 2x$. This choice is equivalent to selecting the origin at the centre of symmetry I or II in the real structure (Fig. 2.3.2.1). Similarly in a three-dimensional $P\bar{1}$ cell, the Patterson will contain peaks at $\langle uvw\rangle$ which can be used to solve for the atom coordinates $\pm\langle 2x, 2y, 2z\rangle$. Solving for the same coordinates by starting from symmetric representations of the same vector will lead to alternate origin choices. For example, use of $\langle 1+u, 1+v, w\rangle$ will lead to translating the origin by $(+\frac{1}{2}, +\frac{1}{2}, 0)$ relative to the solution based on $\langle uvw\rangle$. There are eight distinct inversion centres in $P\bar{1}$, each one of which represents a valid origin choice. Although any choice of origin would be allowable, an inversion centre is convenient because then the structure factors are all real. Typically, one of the vector peaks closest to the Patterson origin is selected to start the solution, usually in the calculated asymmetric unit of the Patterson. Care must be exercised in selecting the same origin for all atomic positions by considering cross-vectors between atoms.

Examine, for example, the *c*-axis Patterson projection of a cuprous chloride azomethane complex ($C_2H_6Cl_2Cu_2N_2$) in $P\bar{1}$ as shown in Fig. 2.3.2.2. The largest Patterson peaks should correspond to vectors arising from Cu ($Z = 29$) and Cl ($Z = 17$) atoms. There will be copper atoms at $\mathbf{x}_{Cu}(x_{Cu}, y_{Cu})$ and $-\mathbf{x}_{Cu}(-x_{Cu}, -y_{Cu})$ as well as chlorine atoms at analogous positions. The interaction matrix is

	$\mathbf{x}_{Cu}$, 29	$\mathbf{x}_{Cl}$, 17	$-\mathbf{x}_{Cu}$, 29	$-\mathbf{x}_{Cl}$, 17
$\mathbf{x}_{Cu}$, 29	0, 841	$\mathbf{x}_{Cu}-\mathbf{x}_{Cl}$, 493	$2\mathbf{x}_{Cu}$, 841	$\mathbf{x}_{Cu}+\mathbf{x}_{Cl}$, 493
$\mathbf{x}_{Cl}$, 17		0, 289	$\mathbf{x}_{Cl}+\mathbf{x}_{Cu}$, 493	$2\mathbf{x}_{Cl}$, 289
$-\mathbf{x}_{Cu}$, 29			0, 841	$\mathbf{x}_{Cu}-\mathbf{x}_{Cl}$, 493
$-\mathbf{x}_{Cl}$, 17				0, 289

Table 2.3.2.1. *Coordinates of Patterson peaks for $C_2H_6Cl_2Cu_2N_2$ projection*

Height	u	v	Number in diagram (Fig. 2.3.2.2)
7	0.33	0.34	I
7	0.18	0.97	II
6	0.16	0.40	III
3	0.49	0.29	IV
3	0.02	0.59	V
2	0.30	0.75	VI
2	0.12	0.79	VII

which shows that the Patterson should contain the following types of vectors:

Position	Weight	Multiplicity	Total weight
$2\mathbf{x}_{\rm Cu}$	841	1	841
$2\mathbf{x}_{\rm Cl}$	289	1	289
$\mathbf{x}_{\rm Cu}-\mathbf{x}_{\rm Cl}$	493	2	986
$\mathbf{x}_{\rm Cu}+\mathbf{x}_{\rm Cl}$	493	2	986

The coordinates of the largest Patterson peaks are given in Table 2.3.2.1 for an asymmetric half of the cell chosen to span $0 \rightarrow \frac{1}{2}$ in u and $0 \rightarrow 1$ in v. Since the three largest peaks are in the same ratio (7:7:6) as the three largest expected vector types (986:986:841), it is reasonable to assume that peak III corresponds to the copper–copper interaction at $2\mathbf{x}_{\rm Cu}$. Hence, $x_{\rm Cu} = 0.08$ and $y_{\rm Cu} = 0.20$. Peaks I and II should be due to the double-weight Cu–Cl vectors at $\mathbf{x}_{\rm Cu} - \mathbf{x}_{\rm Cl}$ and $\mathbf{x}_{\rm Cu} + \mathbf{x}_{\rm Cl}$. Now suppose that peak I is at position $\mathbf{x}_{\rm Cu} + \mathbf{x}_{\rm Cl}$, then $x_{\rm Cl} = 0.25$ and $y_{\rm Cl} = 0.14$. Peak II should now check out as the remaining double-weight Cu–Cl interaction at $\mathbf{x}_{\rm Cu} - \mathbf{x}_{\rm Cl}$. Indeed, $\mathbf{x}_{\rm Cu} - \mathbf{x}_{\rm Cl} = \langle -0.17, 0.06\rangle = -\langle 0.17, -0.06\rangle$ which agrees tolerably well with the position of peak II. The chlorine position also predicts the position of a peak at $2\mathbf{x}_{\rm Cl}$ with weight 289; peak IV confirms the chlorine assignment. In fact, this Patterson can be solved also for the lighter nitrogen- and carbon-atom positions which account for the remainder of the vectors listed in Table 2.3.2.1. However, the simplest way to complete the structure determination is probably to compute a Fourier synthesis using phases calculated from the heavier copper and chlorine positions.

Consider now a real cell with M crystallographic asymmetric units, each of which contains N atoms. Let us define $\mathbf{x}_{mn}$, the position of the nth atom in the mth crystallographic unit, by

$$\mathbf{x}_{mn} = [\boldsymbol{T}_m]\mathbf{x}_{1n} + \mathbf{t}_m,$$

where $[\boldsymbol{T}_m]$ and $\mathbf{t}_m$ are the rotation matrix and translation vector, respectively, for the mth crystallographic symmetry operator. The Patterson of this crystal will contain vector peaks which arise from atoms interacting with other atoms both in the same and in different crystallographic asymmetric units. The set of $(MN)^2$ Patterson vector interactions for this crystal is represented in a matrix in Table 2.3.2.2. Upon dissection of this diagram we see that there are MN origin vectors, $M[(N-1)N]$ vectors from atom interactions with other atoms in the same crystallographic asymmetric unit and $[M(M-1)]N^2$ vectors involving atoms in separate asymmetric units. Often a number of vectors of special significance relating symmetry-equivalent atoms emerge from this milieu of Patterson vectors and such 'Harker vectors' constitute the subject of the next section.

2.3.2.2. Harker sections

Soon after Patterson introduced the F^2 series, Harker (1936) recognized that many types of crystallographic symmetry result in a concentration of vectors at characteristic locations in the Patterson. Specifically, he showed that atoms related by rotation axes produce vectors in characteristic planes of the Patterson, and that atoms related by mirror planes or reflection glide planes produce vectors on characteristic lines. Similarly, noncrystallographic symmetry operators produce analogous concentrations of vectors. Harker showed how special sections through a three-dimensional function could be computed using one- or two-dimensional summations. With the advent of powerful computers, it is usual to calculate a full three-dimensional Patterson synthesis. Nevertheless, 'Harker' planes or lines are often the starting point for a structure determination. It should, however, be noted that non-Harker vectors (those not due to interactions between symmetry-related atoms) can appear by coincidence in a Harker section. Table 2.3.2.3 shows the position in a Patterson of Harker planes and lines produced by all types of crystallographic symmetry operators.

Buerger (1946) noted that Harker sections can be helpful in space-group determination. Concentrations of vectors in appropriate regions of the Patterson should be diagnostic for the presence of some symmetry elements. This is particularly useful where these elements (such as mirror planes) are not directly detected by systematic absences.

Buerger also developed a systematic method of interpreting Harker peaks which he called implication theory [Buerger (1959, Chapter 7)].

Table 2.3.2.2. *Square matrix representation of vector interactions in a Patterson of a crystal with M crystallographic asymmetric units each containing N atoms*

Peak positions $\mathbf{u}_{m1n1,\,m2n2}$ correspond to vectors between the atoms $\mathbf{x}_{m1n1}$ and $\mathbf{x}_{m2n2}$ where $\mathbf{x}_{mn}$ is the nth atom in the mth crystallographic asymmetric unit. The corresponding weights are $w_{n1}w_{n2}$. The outlined blocks I1 and IM represent vector interactions between atoms in the same crystallographic asymmetric units (there are M such blocks). The off-diagonal blocks IIM1 and II1M represent vector interactions between atoms in crystal asymmetric units 1 and M; there are $M(M-1)$ blocks of this type. The significance of diagonal elements of block IIM1 is that they represent Harker-type interactions between symmetry-equivalent atoms (see Section 2.3.2.2).

	$\mathbf{x}_{11}, w_1$	$\mathbf{x}_{12}, w_2$	...	$\mathbf{x}_{1N}, w_N$	...	$\mathbf{x}_{M1}, w_1$	$\mathbf{x}_{M2}, w_2$	...	$\mathbf{x}_{MN}, w_N$
$\mathbf{x}_{11}, w_1$	$0, w_1^2$	$\mathbf{u}_{11,12}, w_1w_2$	...	$\mathbf{u}_{11,1N}, w_1w_N$	...				
$\mathbf{x}_{12}, w_2$		$0, w_2^2$	...	$\mathbf{u}_{12,1N}, w_2w_N$					
⋮		⋱		⋮	...				
$\mathbf{x}_{1N}, w_N$				$0, w_N^2$					
	Block I1					Block II1M			
⋮	⋮				⋱		⋮		
$\mathbf{x}_{M1}, w_1$	$\mathbf{u}_{M1,11}, w_1^2$	$\mathbf{u}_{M1,12}, w_1w_2$							
$\mathbf{x}_{M2}, w_2$	$\mathbf{u}_{M2,11}, w_2w_1$	$\mathbf{u}_{M2,12}, w_2^2$							
⋮		⋱			...				
$\mathbf{x}_{MN}, w_N$				$\mathbf{u}_{MN,1N}, w_N^2$					
	Block IIM1					Block IM			

Table 2.3.2.3. *Position of Harker sections within a Patterson*

Symmetry element	Form of $P(x, y, z)$
(*a*) Harker planes	
Axes parallel to the *b* axis:	
(i) 2	$P(x, 0, z)$
(ii) 2_1	$P(x, \frac{1}{2}, z)$
Axes parallel to the *c* axis:	
(i) 2, 3, $\bar{3}$, 4, $\bar{4}$, 6, $\bar{6}$	$P(x, y, 0)$
(ii) 2_1, 4_2, 6_3	$P(x, y, \frac{1}{2})$
(iii) 3_1, 3_2, 6_2, 6_4	$P(x, y, \frac{1}{3})$
(iv) 4_1, 4_3	$P(x, y, \frac{1}{4})$
(v) 6_1, 6_5	$P(x, y, \frac{1}{6})$
(*b*) Harker lines	
Planes perpendicular to the *b* axis:	
(i) Reflection planes	$P(0, y, 0)$
(ii) Glide plane, glide $= \frac{1}{2}a$	$P(\frac{1}{2}, y, 0)$
(iii) Glide plane, glide $= \frac{1}{2}c$	$P(0, y, \frac{1}{2})$
(iv) Glide plane, glide $= \frac{1}{2}(a + c)$	$P(\frac{1}{2}, y, \frac{1}{2})$
(v) Glide plane, glide $= \frac{1}{4}(a + c)$	$P(\frac{1}{4}, y, \frac{1}{4})$
(vi) Glide plane, glide $= \frac{1}{4}(3a + c)$	$P(\frac{3}{4}, y, \frac{1}{4})$
(*c*) Special Harker planes	
Axes parallel to or containing body diagonal (111), valid for cubic space groups only:	
	Equation of plane $lx + my + nz - p = 0$
(i) 3	$l = m = n = \cos 54.73561^\circ = 0.57735$, $p = 0$
(ii) 3_1	$l = m = n = \cos 54.73561^\circ = 0.57735$, $p = \sqrt{3}/3$
Rhombohedral threefold axes produce analogous Harker planes whose description will depend on the interaxial angle.	

2.3.2.3. *Finding heavy atoms*

The previous two sections have developed some of the useful mechanics for interpreting Pattersons. In this section, we will consider finding heavy-atom positions, in the presence of numerous light atoms, from Patterson maps. The feasibility of structure solution by the heavy-atom method depends on a number of factors which include the relative size of the heavy atom and the extent and quality of the data. A useful rule of thumb is that the ratio

$$r = \frac{\sum_{\text{heavy}} Z^2}{\sum_{\text{light}} Z^2}$$

should be near unity if the heavy atom is to provide useful starting phase information (Z is the atomic number of an atom). The condition that $r > 1$ normally guarantees interpretability of the Patterson function in terms of the heavy-atom positions. This 'rule', arising from the work of Luzzati (1953), Woolfson (1956), Sim (1961) and others, is not inviolable; many ambitious determinations have been accomplished *via* the heavy-atom method for which r was well below 1.0. An outstanding example is vitamin B_{12} with formula $C_{62}H_{88}CoO_{14}P$, which gave an $r = 0.14$ for the cobalt atom alone (Hodgkin *et al.*, 1957). One factor contributing to the success of such a determination is that the relative scattering power of Co is enhanced for higher scattering angles. Thus, the ratio, r, provides a conservative estimate. If the value of r is well above 1.0, the initial easier interpretation of the Patterson will come at the expense of poorly defined parameters of the lighter atoms.

A general strategy for determining heavy atoms from the Patterson usually involves the following steps.

(1) List the number and type of atoms in the cell.

(2) Construct the interaction matrix for the heaviest atoms to predict the positions and weights of the largest Patterson vectors. Group recurrent vectors and notice vectors with special properties, such as Harker vectors.

(3) Compute the Patterson using any desired modifications. Placing the map on an absolute scale [$P(000) = \sum Z^2$] is convenient but not necessary.

(4) Examine Harker sections and derive trial atom coordinates from vector positions.

(5) Check the trial coordinates using other vectors in the predicted set. Correlate enantiomorphic choice and origin choice for independent sites.

(6) Include the next-heaviest atoms in the interpretation if possible. In particular, use the cross-vectors with the heaviest atoms.

(7) Use the best heavy-atom model to initiate phasing.

Detailed and instructive examples of using Pattersons to find heavy-atom positions are found in almost every textbook on crystal structure analysis [see, for example, Buerger (1959), Lipson & Cochran (1966) and Stout & Jensen (1968)].

The determination of the crystal structure of cholesteryl iodide by Carlisle & Crowfoot (1945) provides an example of using the Patterson function to locate heavy atoms. There were two molecules, each of formula $C_{27}H_{45}I$, in the $P2_1$ unit cell. The ratio $r = 2.8$ is clearly well over the optimal value of unity. The $P(x, z)$ Patterson projection showed one dominant peak at $\langle 0.434, 0.084 \rangle$ in the asymmetric unit. The equivalent positions for $P2_1$ require that an iodine atom at x_I, y_I, z_I generates another at $-x_I, \frac{1}{2} + y_I, z_I$ and thus produces a Patterson peak at $\langle 2x_I, \frac{1}{2}, 2z_I \rangle$. The iodine position was therefore determined as 0.217, 0.042. The y coordinate of the iodine is arbitrary for $P2_1$ yet the value of $y_I = 0.25$ is convenient, since an inversion centre in the two-atom iodine structure is then exactly at the origin, making all calculated phases 0 or π. Although the presence of this extra symmetry caused some initial difficulties in the interpretation of the steroid backbone, Carlisle and Crowfoot successfully separated the enantiomorphic images. Owing to the presence of the perhaps *too heavy* iodine atom, however, the structure of the carbon skeleton could not be defined very precisely. Nevertheless, all critical stereochemical details were adequately illuminated by this determination. In the cholesteryl iodide example, a number of different yet equivalent origins could have been selected. Alternative origin choices include all combinations of $x \pm \frac{1}{2}$ and $z \pm \frac{1}{2}$.

A further example of using the Patterson to find heavy atoms will be provided in Section 2.3.5.2 on solving for heavy atoms in the presence of noncrystallographic symmetry.

2.3.2.4. *Superposition methods. Image detection*

As early as 1939, Wrinch (1939) showed that it was possible, in principle, to recover a fundamental set of points from the vector map of that set. Unlike the Harker–Buerger implication theory (Buerger, 1946), the method that Wrinch suggested was capable of using all the vectors in a three-dimensional set, not those restricted to special lines or sections. Wrinch's ideas were developed for vector sets of points, however, and could not be directly applied to real, heavily overlapped Pattersons of a complex structure. These ideas seem to have lain dormant until the early 1950s when a number of independent investigators developed superposition methods (Beevers & Robertson, 1950; Clastre & Gay, 1950; Garrido, 1950*a*; Buerger, 1950*a*).

A Patterson can be considered as a sum of images of a molecule as seen, in turn, for each atom placed on the origin (Fig. 2.3.2.3). Thus, the deconvolution of a Patterson could proceed by superimposing each image of the molecule obtained onto the others by translating the Patterson origin to each imaging atom.

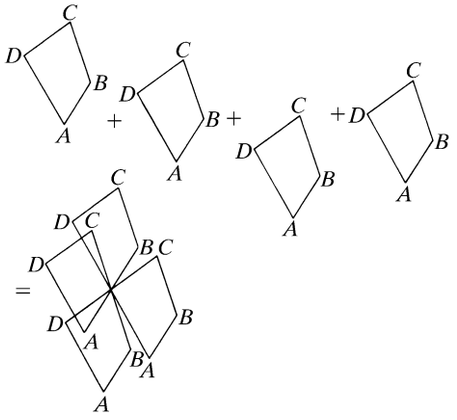


Fig. 2.3.2.3. Atoms *ABCD*, arranged as a quadrilateral, generate a Patterson which is the sum of the images of the quadrilateral when each atom is placed on the origin in turn.

For instance, let us take a molecule consisting of four atoms *ABCD* arranged in the form of a quadrilateral (Fig. 2.3.2.3). Then the Patterson consists of the images of four identical quadrilaterals with atoms *A*, *B*, *C* and *D* placed on the origin in turn. The Pattersons can then be deconvoluted by superimposing two of these Pattersons after translating these (without rotation) by, for instance, the vector **AB**. A further improvement could be obtained by superimposing a third Patterson translated by **AC**. This would have the additional advantage in that *ABC* is a noncentrosymmetric arrangement and, therefore, selects the enantiomorph corresponding to the hand of the atomic arrangement *ABC* [*cf.* Buerger (1951, 1959)].

A basic problem is that knowledge of the vectors **AB** and **AC** also implies some knowledge of the structure at a time when the structure is not yet known. In practice 'good-looking' peaks, estimated to be single peaks by assessing the absolute scale of the structure amplitudes with Wilson statistics, can be assumed to be the result of single interatomic vectors within a molecule. Superposition can then proceed and the result can be inspected for reasonable chemical sense. As many apparently single peaks can be tried systematically using a computer, this technique is useful for selecting and testing a series of reasonable Patterson interpretations (Jacobson *et al.*, 1961).

Three major methods have been used for the detection of molecular images of superimposed Pattersons. These are the sum, product and minimum 'image seeking' functions (Raman & Lipscomb, 1961). The concept of the sum function is to add the images where they superimpose, whereas elsewhere the summed Pattersons will have a lower value owing to lack of image superposition. Therefore, the sum function determines the average Patterson density for superimposed images, and is represented analytically as

$$S(\mathbf{x}) = \sum_{i=1}^{N} P(\mathbf{x} + \mathbf{u}_i),$$

where $S(\mathbf{x})$ is the sum function at $\mathbf{x}$ given by the superposition of the ith Patterson translated by $\mathbf{u}_i$, or

$$S(\mathbf{x}) = \sum_{\mathbf{h}} \left\{ F_{\mathbf{h}}^2 \exp(2\pi i \mathbf{h} \cdot \mathbf{x}) \left[\sum_{i=1}^{N} \exp(2\pi i \mathbf{h} \cdot \mathbf{u}_i) \right] \right\}.$$

Setting

$$m \exp(i\varphi_{\mathbf{h}}) = \sum_{i=1}^{N} \exp(2\pi i \mathbf{h} \cdot \mathbf{u}_i)$$

(m and $\varphi_{\mathbf{h}}$ can be calculated from the translational vectors used for the superposition),

$$S(\mathbf{x}) = \sum_{\mathbf{h}} F_{\mathbf{h}}^2 m \exp(2\pi i \mathbf{h} \cdot \mathbf{x} + \varphi_{\mathbf{h}}).$$

Thus, the sum function is equivalent to a weighted 'heavy atom' method based on the known atoms assumed by the superposition translation vectors.

The product function is somewhat more vigorous in that the images are enhanced by the product. If an image is superimposed on no image, then the product should be zero.

The product function can be expressed as

$$Pr(\mathbf{x}) = \prod_{i=1}^{N} P(\mathbf{x} + \mathbf{u}_i).$$

When $N = 2$ ($\mathbf{h}$ and $\mathbf{p}$ are sets of Miller indices),

$$Pr(\mathbf{x}) = \sum_{\mathbf{h}} \sum_{\mathbf{p}} F_{\mathbf{h}}^2 F_{\mathbf{p}}^2 \exp[2\pi i(\mathbf{h} + \mathbf{p}) \cdot \mathbf{x}] \times \exp[2\pi i(\mathbf{h} \cdot \mathbf{u}_i + \mathbf{p} \cdot \mathbf{u}_i)].$$

Successive superpositions using the product functions will quickly be dominated by a few terms with very large coefficients.

Finally, the minimum function is a clever invention of Buerger (Buerger, 1950*b*, 1951, 1953*a*,*b*,*c*; Taylor, 1953; Rogers, 1951). If a superposition is correct then each Patterson must represent an image of the structure. Whenever there are two or more images that intersect in the Patterson, the Patterson density will be greater than a single image. When two different images are superimposed, it is a reasonable hope that at least one of these is a single image. Thus by taking the value of that Patterson which is the minimum, it should be possible to select a single image and eliminate noise from interfering images as far as possible. Although the minimum function is perhaps the most powerful algorithm for image selection of well sharpened Pattersons, it is not readily amenable to Fourier representation.

The minimum function was conceived on the basis of selecting positive images on a near-zero background. If it were desired to select negative images [*e.g.* the $(\mathbf{F}_1 - \mathbf{F}_2)^2$ correlation function discussed in Section 2.3.3.4], then it would be necessary to use a maximum function. In fact, normally, an image has finite volume with varying density. Thus, some modification of the minimum function is necessary in those cases where the image is large compared to the volume of the unit cell, as in low-resolution protein structures (Rossmann, 1961*b*). Nordman (1966) used the average of the Patterson values of the lowest 10 to 20 per cent of the vectors in comparing Pattersons with hypothetical point Pattersons. A similar criterion was used by High & Kraut (1966).

Image-seeking methods using Patterson superposition have been used extensively (Beevers & Robertson, 1950; Garrido, 1950*b*; Robertson, 1951). For a review the reader is referred to *Vector Space* (Buerger, 1959) and a paper by Fridrichsons & Mathieson (1962). However, with the advent of computerized direct methods (see Chapter 2.2), such techniques are no longer popular. Nevertheless, they provide the conceptual framework for the rotation and translation functions (see Sections 2.3.6 and 2.3.7).

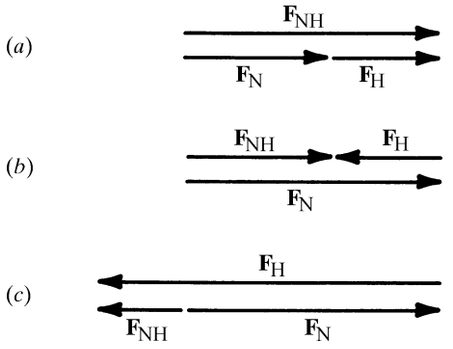


Fig. 2.3.3.1. Three different cases which can occur in the relation of the native, $\mathbf{F}_N$, and heavy-atom derivative, $\mathbf{F}_{NH}$, structure factors for centrosymmetric reflections. $\mathbf{F}_N$ is assumed to have a phase of 0, although analogous diagrams could be drawn when $\mathbf{F}_N$ has a phase of π. The crossover situation in (*c*) is clearly rare if the heavy-atom substitution is small compared to the native molecule, and can in general be neglected.

2.3.2.5. Systematic computerized Patterson vector-search procedures. Looking for rigid bodies

The power of the modern digital computer has enabled rapid access to the large number of Patterson density values which can serve as a lookup table for systematic vector-search procedures. In the late 1950s, investigators began to use systematic searches for the placement of single atoms, of known chemical groups or fragments and of entire known structures. Methods for locating single atoms were developed and called variously: vector verification (Mighell & Jacobson, 1963), symmetry minimum function (Kraut, 1961; Simpson *et al*., 1965; Corfield & Rosenstein, 1966) and consistency functions (Hamilton, 1965). Patterson superposition techniques using stored function values were often used to image the structure from the known portion. In such single-site search procedures, single atoms are placed at all possible positions in a crystal, using a search grid of the same fineness as for the stored Patterson function, preferably about one-third of the resolution of the Patterson map. Solutions are gauged to be acceptable if all expected vectors due to symmetry-related atoms are observed above a specified threshold in the Patterson.

Systematic computerized Patterson search procedures for orienting and positioning known molecular fragments were also developed in the early 1960s. These hierarchical procedures rely on first using the 'self'-vectors which depend only on the orientation of a molecular fragment. A search for the position of the fragment relative to the crystal symmetry elements then uses the cross-vectors between molecules (see Sections 2.3.6 and 2.3.7). Nordman constructed a weighted point representation of the predicted vector set for a fragment (Nordman & Nakatsu, 1963; Nordman, 1966) and successfully solved the structure of a number of complex alkaloids. Huber (1965) used the convolution molecule method of Hoppe (1957*a*) in three dimensions to solve a number of natural-product structures, including steroids. Various program systems have used Patterson search methods operating in real space to solve complex structures (Braun *et al*., 1969; Egert, 1983).

Others have used reciprocal-space procedures for locating known fragments. Tollin & Cochran (1964) developed a procedure for determining the orientation of planar groups by searching for origin-containing planes of high density in the Patterson. General procedures using reciprocal-space representations for determining rotation and translation parameters have been developed and will be described in Sections 2.3.6 and 2.3.7, respectively.

Although as many functions have been used to detect solutions in these Patterson search procedures as there are programs, most rely on some variation of the sum, product and minimum functions (Section 2.3.2.4). The quality of the stored Patterson density representation also varies widely, but it is now common to use 16 or more bits for single density values. Treatment of vector overlap is handled differently by different investigators and the choice will depend on the degree of overlapping (Nordman & Schilling, 1970; Nordman, 1972). General Gaussian multiplicity corrections can be employed to treat coincidental overlap of independent vectors in general positions and overlap which occurs for symmetric peaks in the vicinity of a special position or mirror plane in the Patterson (G. Kamer, S. Ramakumar & P. Argos, unpublished results; Rossmann *et al*., 1972).

2.3.3. Isomorphous replacement difference Pattersons

2.3.3.1. Introduction

One of the initial stages in the application of the isomorphous replacement method is the determination of heavy-atom positions. Indeed, this step of a structure determination can often be the most challenging. Not only may the number of heavy-atom sites be unknown, and have incomplete substitution, but the various isomorphous compounds may also lack isomorphism. To compound these problems, the error in the measurement of the isomorphous difference in structure amplitudes is often comparable to the differences themselves. Clearly, therefore, the ease with which a particular problem can be solved is closely correlated with the quality of the data-measuring procedure.

The isomorphous replacement method was used incidentally by Bragg in the solution of NaCl and KCl. It was later formalized by J. M. Robertson in the analysis of phthalocyanine where the coordination centre could be Pt, Ni and other metals (Robertson, 1935, 1936; Robertson & Woodward, 1937). In this and similar cases, there was no difficulty in finding the heavy-atom positions. Not only were the heavy atoms frequently in special positions, but they dominated the total scattering effect. It was not until Perutz and his colleagues (Green *et al*., 1954; Bragg & Perutz, 1954) applied the technique to the solution of haemoglobin, a protein of 68 000 Da, that it was necessary to consider methods for detecting heavy atoms. The effect of a single heavy atom, even uranium, can only have a very marginal effect on the structure amplitudes of a crystalline macromolecule. Hence, techniques had to be developed which were dependent on the difference of the isomorphous structure amplitudes rather than on the solution of the Patterson of the heavy-atom-derivative compound on its own.

2.3.3.2. Finding heavy atoms with centrosymmetric projections

Phases in a centrosymmetric projection will be 0 or π if the origin is chosen at the centre of symmetry. Hence, the native structure factor, $\mathbf{F}_N$, and the heavy-atom-derivative structure factor, $\mathbf{F}_{NH}$, will be collinear. It follows that the structure amplitude, $|\mathbf{F}_H|$, of the heavy atoms alone in the cell will be given by

$$|\mathbf{F}_H| = |(|\mathbf{F}_{NH}| \pm |\mathbf{F}_N|)| + \varepsilon,$$

where ε is the error on the parenthetic sum or difference. Three different cases may arise (Fig. 2.3.3.1). Since the situation shown in Fig. 2.3.3.1(*c*) is rare, in general

$$|\mathbf{F}_H|^2 \simeq (|\mathbf{F}_{NH}| - |\mathbf{F}_N|)^2. \qquad (2.3.3.1)$$

Thus, a Patterson computed with the square of the differences between the native and derivative structure amplitudes of a centrosymmetric projection will approximate to a Patterson of the heavy atoms alone.

The approximation (2.3.3.1) is valid if the heavy-atom substitution is small enough to make $|\mathbf{F}_H| \ll |\mathbf{F}_{NH}|$ for most reflections, but sufficiently large to make $\varepsilon \ll (|\mathbf{F}_{NH}| - |\mathbf{F}_N|)^2$. It is also

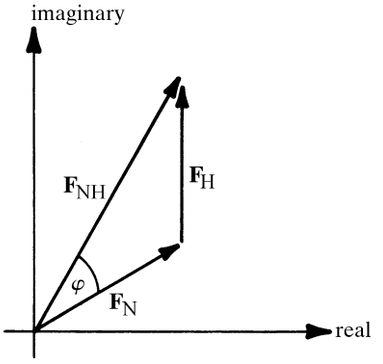


Fig. 2.3.3.2. Vector triangle showing the relationship between $\mathbf{F}_N$, $\mathbf{F}_{NH}$ and $\mathbf{F}_H$, where $\mathbf{F}_{NH} = \mathbf{F}_N + \mathbf{F}_H$.

assumed that the native and heavy-atom-derivative data have been placed on the same relative scale. Hence, the relation (2.3.3.1) should be rewritten as

$$|\mathbf{F}_H|^2 \simeq (|\mathbf{F}_{NH}| - k|\mathbf{F}_N|)^2,$$

where k is an experimentally determined scale factor (see Section 2.3.3.7). Uncertainty in the determination of k will contribute further to ε, albeit in a systematic manner.

Centrosymmetric projections were used extensively for the determination of heavy-atom sites in early work on proteins such as haemoglobin (Green *et al.*, 1954), myoglobin (Bluhm *et al.*, 1958) and lysozyme (Poljak, 1963). However, with the advent of faster data-collecting techniques, low-resolution (*e.g.* a 5 Å limit) three-dimensional data are to be preferred for calculating difference Pattersons. For noncentrosymmetric reflections, the approximation (2.3.3.1) is still valid but less exact (Section 2.3.3.3). However, the larger number of three-dimensional differences compared to projection differences will enhance the signal of the real Patterson peaks relative to the noise. If there are N terms in the Patterson synthesis, then the peak-to-noise ratio will be proportionally $\sqrt{N}$ and $1/\varepsilon$. With the subscripts 2 and 3 representing two- and three-dimensional syntheses, respectively, the latter will be more powerful than the former whenever

$$\frac{\sqrt{N_3}}{\varepsilon_3} > \frac{\sqrt{N_2}}{\varepsilon_2}.$$

Now, as $\varepsilon_3 \simeq \sqrt{2}\varepsilon_2$, it follows that N_3 must be greater than $2N_2$ if the three-dimensional noncentrosymmetric computation is to be more powerful. This condition must almost invariably be true.

2.3.3.3. Finding heavy atoms with three-dimensional methods

A Patterson of a native biomacromolecular structure (coefficients F_N^2) can be considered as being, at least approximately, a vector map of all the light atoms (carbons, nitrogens, oxygens, some sulfurs, and also phosphorus for nucleic acids) other than hydrogen atoms. These interactions will be designated as LL. Similarly, a Patterson of the heavy-atom derivative will contain $HH + HL + LL$ interactions, where H represents the heavy atoms. Thus, a true difference Patterson, with coefficients $F_{NH}^2 - F_N^2$, will contain only the interactions $HH + HL$. In general, the carpet of HL vectors completely dominates the HH vectors except for very small proteins such as insulin (Adams *et al.*, 1969). Therefore, it would be preferable to compute a Patterson containing only HH interactions in order to interpret the map in terms of specific heavy-atom sites.

Blow (1958) and Rossmann (1960) showed that a Patterson with $(|\mathbf{F}_{NH}| - |\mathbf{F}_N|)^2$ coefficients approximated to a Patterson containing only HH vectors. If the phase angle between $\mathbf{F}_N$ and $\mathbf{F}_{NH}$ is φ (Fig. 2.3.3.2), then

$$|\mathbf{F}_H|^2 = |\mathbf{F}_N|^2 + |\mathbf{F}_{NH}|^2 - 2|\mathbf{F}_N||\mathbf{F}_{NH}|\cos\varphi.$$

In general, however, $|\mathbf{F}_H| \ll |\mathbf{F}_N|$. Hence, φ is small and

$$|\mathbf{F}_H|^2 \simeq (|\mathbf{F}_{NH}| - |\mathbf{F}_N|)^2,$$

which is the same relation as (2.3.3.1) for centrosymmetric approximations. Since the direction of $\mathbf{F}_H$ is random compared to $\mathbf{F}_N$, the root-mean-square projected length of $\mathbf{F}_H$ onto $\mathbf{F}_N$ will be $\mathbf{F}_H/\sqrt{2}$. Thus it follows that a better approximation is

$$|\mathbf{F}_H|^2 \simeq \sqrt{2}(|\mathbf{F}_{NH}| - |\mathbf{F}_N|)^2, \qquad (2.3.3.2)$$

which accounts for the assumption (Section 2.3.3.2) that $\varepsilon_3 = \sqrt{2}\varepsilon_2$. The almost universal method for the initial determination of major heavy-atom sites in an isomorphous derivative utilizes a Patterson with $(|\mathbf{F}_{NH}| - |\mathbf{F}_N|)^2$ coefficients. Approximation (2.3.3.2) is also the basis for the refinement of heavy-atom parameters in a single isomorphous replacement pair (Rossmann, 1960; Cullis *et al.*, 1962; Terwilliger & Eisenberg, 1983).

2.3.3.4. Correlation functions

In the most general case of a triclinic space group, it will be necessary to select an origin arbitrarily, usually coincident with a heavy atom. All other heavy atoms (and subsequently also the macromolecular atoms) will be referred to this reference atom. However, the choice of an origin will be independent in the interpretation of each derivative's difference Patterson. It will then be necessary to correlate the various, arbitrarily chosen, origins. The same problem occurs in space groups lacking symmetry axes perpendicular to the primary rotation axis (*e.g.* $P2_1$, $P6$ *etc.*), although only one coordinate, namely parallel to the unique rotation axis, will require correlation. This problem gave rise to some concern in the 1950s. Bragg (1958), Blow (1958), Perutz (1956), Hoppe (1959) and Bodo *et al.* (1959) developed a variety of techniques, none of which were entirely satisfactory. Rossmann (1960) proposed the $(\mathbf{F}_{NH1} - \mathbf{F}_{NH2})^2$ synthesis and applied it successfully to the heavy-atom determination of horse haemoglobin. This function gives positive peaks ($H1 \cdot H1$) at the end of vectors between the heavy-atom sites in the first compound, positive peaks ($H2 \cdot H2$) between the sites in the second compound, and negative peaks between sites in the first and second compound (Fig. 2.3.3.3). It is thus the negative peaks which provide the necessary correlation. The function is unique in that it is a Patterson containing significant information in both positive and negative peaks. Steinrauf (1963) suggested using the coefficients $(|\mathbf{F}_{NH1}| - |\mathbf{F}_N|) \cdot (|\mathbf{F}_{NH2}| - |\mathbf{F}_N|)$ in order to eliminate the positive $H1 \cdot H1$ and $H2 \cdot H2$ vectors.

Although the problem of correlation was a serious concern in the early structural determination of proteins during the late 1950s and early 1960s, the problem has now been bypassed. Blow & Rossmann (1961) and Kartha (1961) independently showed that it was possible to compute usable phases from a single isomorphous replacement (SIR) derivative. This contradicted the previously accepted notion that it was necessary to have at least two isomorphous derivatives to be able to determine a noncentrosymmetric reflection's phase (Harker, 1956). Hence, currently, the procedure used to correlate origins in different derivatives is to compute SIR phases from the first compound and apply them to a difference electron-density map of the second heavy-atom derivative. Thus, the origin of the second

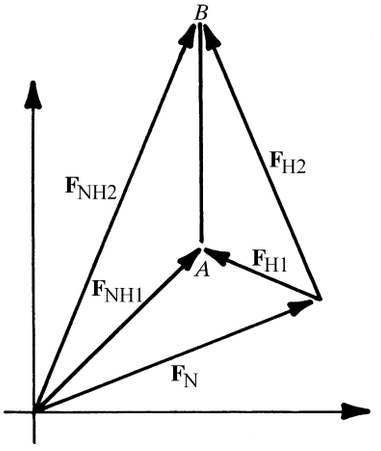


Fig. 2.3.3.3. A Patterson with coefficients $(\mathbf{F}_{NH1} - \mathbf{F}_{NH2})^2$ will be equivalent to a Patterson whose coefficients are $(AB)^2$. However, $AB = -\mathbf{F}_{H1} + \mathbf{F}_{H2}$. Thus, a Patterson with $(AB)^2$ coefficients is equivalent to having negative atomic substitutions in compound 1 and positive substitutions in compound 2, or *vice versa*. Therefore, the Patterson will contain positive peaks for vectors of the type $H1 \cdot H1$ and $H2 \cdot H2$, but negative vector peaks for vectors of type $H1 \cdot H2$.

derivative will be referred to the arbitrarily chosen origin of the first compound. More important, however, the interpretation of such a 'feedback' difference Fourier is easier than that of a difference Patterson. Hence, once one heavy-atom derivative has been solved for its heavy-atom sites, the solution of other derivatives is almost assured. This concept is examined more closely in the following section.

2.3.3.5. Interpretation of isomorphous difference Pattersons

Difference Pattersons have usually been manually interpreted in terms of point atoms. In more complex situations, such as crystalline viruses, a systematic approach may be necessary to analyse the Patterson. That is especially true when the structure contains noncrystallographic symmetry (Argos & Rossmann, 1976). Such methods are in principle dependent on the comparison of the observed Patterson, $P_1(\mathbf{x})$, with a calculated Patterson, $P_2(\mathbf{x})$. A criterion, C_P, based on the sum of the Patterson densities at all test vectors within the unit-cell volume V, would be

$$C_P = \int_V P_1(\mathbf{x}) \cdot P_2(\mathbf{x})\, d\mathbf{x}.$$

C_P can be evaluated for all reasonable heavy-atom distributions. Each different set of trial sites corresponds to a different P_2 Patterson. It is then easily shown that

$$C_P = \sum_{\mathbf{h}} \Delta_{\mathbf{h}}^2 E_{\mathbf{h}}^2,$$

where the sum is taken over all $\mathbf{h}$ reflections in reciprocal space, $\Delta_{\mathbf{h}}^2$ are the observed differences and $E_{\mathbf{h}}$ are the structure factors of the trial point Patterson. (The symbol E is used here because of its close relation to normalized structure factors.)

Let there be n noncrystallographic asymmetric units within the crystallographic asymmetric unit and m crystallographic asymmetric units within the crystal unit cell. Then there are L symmetry-related heavy-atom sites where $L = nm$. Let the scattering contribution of the ith site have a_i and b_i real and imaginary structure-factor components with respect to an arbitrary origin. Hence, for reflection $\mathbf{h}$

$$E_{\mathbf{h}}^2 \left(\sum_L a_{\mathbf{h}i}\right)^2 + \left(\sum_L b_{\mathbf{h}i}\right)^2 = L + \sum_{i\neq j}^{N}\sum^{N} (a_{\mathbf{h}i}a_{\mathbf{h}j} + b_{\mathbf{h}i}b_{\mathbf{h}j}).$$

Therefore,

$$C_P = \sum_{\mathbf{h}} \Delta_{\mathbf{h}}^2 \left[L + 2 \sum_{i\neq j}\sum (a_{\mathbf{h}i}a_{\mathbf{h}j} + b_{\mathbf{h}i}b_{\mathbf{h}j}) \right].$$

But $\sum_{\mathbf{h}} \Delta_{\mathbf{h}}^2$ must be independent of the number, L, of heavy-atom sites per cell. Thus the criterion can be rewritten as

$$C'_P = \sum_{\mathbf{h}} \Delta_{\mathbf{h}}^2 \left[\sum_{i\neq j}\sum (a_{\mathbf{h}i}a_{\mathbf{h}j} + b_{\mathbf{h}i}b_{\mathbf{h}j}) \right]. \quad (2.3.3.3)$$

More generally, if some sites have already been tentatively determined, and if these sites give rise to the structure-factor components $A_{\mathbf{h}}$ and $B_{\mathbf{h}}$, then

$$E_{\mathbf{h}}^2 = \left(A_{\mathbf{h}} + \sum_N a_{\mathbf{h}i}\right)^2 + \left(B_{\mathbf{h}} + \sum_N b_{\mathbf{h}i}\right)^2. \quad (2.3.3.4)$$

Following the same procedure as above, it follows that

$$C'_P = \sum_{\mathbf{h}} \Delta_{\mathbf{h}}^2 \left[(A_{\mathbf{h}}a_{\mathbf{h}} + B_{\mathbf{h}}b_{\mathbf{h}}) + \sum_{i\neq j}\sum (a_{\mathbf{h}i}a_{\mathbf{h}j} + b_{\mathbf{h}i}b_{\mathbf{h}j}) \right], \quad (2.3.3.5)$$

where $a_{\mathbf{h}} = \sum_{i=1}^{L} a_{\mathbf{h}i}$ and $b_{\mathbf{h}} = \sum_{i=1}^{L} b_{\mathbf{h}i}$.

Expression (2.3.3.5) will now be compared with the 'feedback' method (Dickerson *et al.*, 1967, 1968) of verifying heavy-atom sites using SIR phasing. Inspection of Fig. 2.3.3.4 shows that the native phase, α, will be determined as $\alpha = \varphi + \pi$ (φ is the structure-factor phase corresponding to the presumed heavy-atom positions) when $|\mathbf{F}_N| > |\mathbf{F}_H|$ and $\alpha = \varphi$ when $|\mathbf{F}_N| \ll |\mathbf{F}_H|$. Thus, an SIR difference electron density, $\Delta\rho(\mathbf{x})$, can be synthesized by the Fourier summation

$$\begin{aligned}\Delta\rho(\mathbf{x}) =& \frac{1}{V}\sum m(|\mathbf{F}_{NH}| - |\mathbf{F}_N|)\cos(2\pi\mathbf{h}\cdot\mathbf{x} - \varphi_{\mathbf{h}}) \\ & \text{from terms with } \Delta_{\mathbf{h}} = |\mathbf{F}_{NH}| - |\mathbf{F}_N| > 0 \\ & + \frac{1}{V}\sum m(|\mathbf{F}_{NH}| - |\mathbf{F}_N|)\cos(2\pi\mathbf{h}\cdot\mathbf{x} - \varphi_{\mathbf{h}} - \pi) \\ & \text{from terms with } \Delta_{\mathbf{h}} < 0 \\ =& \frac{1}{V}\sum m|\Delta_{\mathbf{h}}|\cos(2\pi\mathbf{h}\cdot\mathbf{x} - \varphi_{\mathbf{h}}),\end{aligned}$$

where m is a figure of merit of the phase reliability (Blow & Crick, 1959; Dickerson *et al.*, 1961). Now,

$$\mathbf{F}_{\mathbf{h}} = A_{\mathbf{h}} + iB_{\mathbf{h}} = F_H \cos\varphi_{\mathbf{h}} + iF_H \sin\varphi_{\mathbf{h}},$$

where $A_{\mathbf{h}}$ and $B_{\mathbf{h}}$ are the real and imaginary components of the presumed heavy-atom sites. Therefore,

$$\Delta\rho(\mathbf{x}) = \frac{1}{V}\sum \frac{m|\Delta_{\mathbf{h}}|}{|\mathbf{F}_H|}(A_{\mathbf{h}}\cos 2\pi\mathbf{h}\cdot\mathbf{x} + B_{\mathbf{h}}\sin 2\pi\mathbf{h}\cdot\mathbf{x}).$$

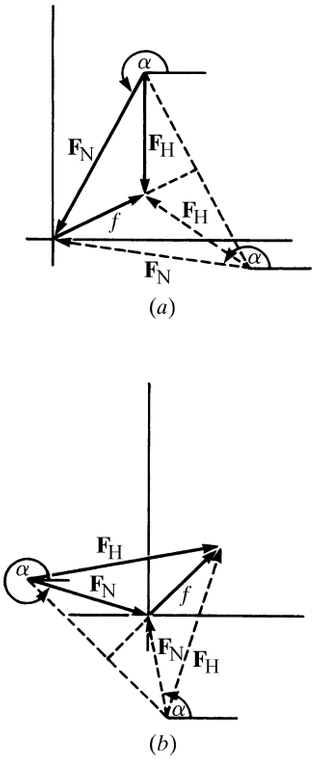


Fig. 2.3.3.4. The phase α of the native compound (structure factor $\mathbf{F}_N$) is determined either as being equal to, or 180° out of phase with, the presumed heavy-atom contribution when only a single isomorphous compound is available. In (*a*) is shown the case when $|\mathbf{F}_N| > |\mathbf{F}_{NH}|$ and $\alpha \simeq \varphi + \pi$. In (*b*) is shown the case when $|\mathbf{F}_N| < |\mathbf{F}_{NH}|$ and $\alpha = \varphi$, where φ is the phase of the heavy-atom structure factor $\mathbf{F}_H$.

If this SIR difference electron-density map shows significant peaks at sites related by noncrystallographic symmetry, then those sites will be at the position of a further set of heavy atoms. Hence, a suitable criterion for finding heavy-atom sites is

$$C_{\rm SIR} = \sum_{j=1}^{n} \Delta\rho(\mathbf{x}_j),$$

or by substitution

$$C_{\rm SIR} = \sum_{j=1}^{n} \frac{1}{V} \sum_{\mathbf{h}} \frac{m|\Delta_{\mathbf{h}}|}{|\mathbf{F}_H|} (A_{\mathbf{h}} \cos 2\pi\mathbf{h}\cdot\mathbf{x}_j + B_{\mathbf{h}} \sin 2\pi\mathbf{h}\cdot\mathbf{x}_j).$$

But

$$a_{\mathbf{h}} = \sum_{j=1}^{n} \cos 2\pi\mathbf{h}\cdot\mathbf{x}_j \quad \text{and} \quad b_{\mathbf{h}} = \sum_{j=1}^{n} \sin 2\pi\mathbf{h}\cdot\mathbf{x}_j.$$

Therefore,

$$C_{\rm SIR} = \frac{1}{V} \sum_{\mathbf{h}} \frac{m|\Delta_{\mathbf{h}}|}{|\mathbf{F}_H|} (A_{\mathbf{h}} a_{\mathbf{h}} + B_{\mathbf{h}} b_{\mathbf{h}}). \qquad (2.3.3.6)$$

This expression is similar to (2.3.3.5) derived by consideration of a Patterson search. It differs from (2.3.3.5) in two respects: the Fourier coefficients are different and expression (2.3.3.6) is lacking a second term. Now the figure of merit m will be small whenever $|\mathbf{F}_H|$ is small as the SIR phase cannot be determined well under those conditions. Hence, effectively, the coefficients are a function of $|\Delta_{\mathbf{h}}|$, and the coefficients of the functions

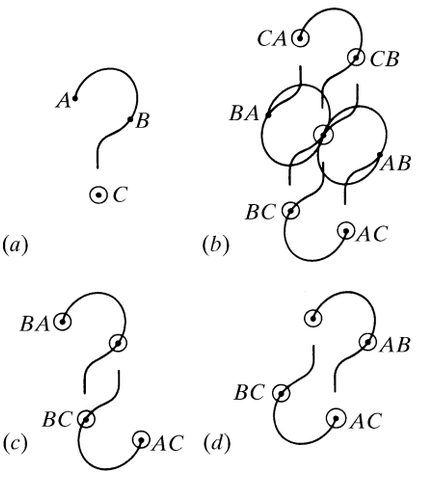


Fig. 2.3.3.5. Let (*a*) be the original structure which contains three heavy atoms *ABC* in a noncentrosymmetric configuration. Then a Fourier summation, with $(F_{NH}^2 - F_N^2)$ coefficients, will give the Patterson shown in (*b*). Displacement of the Patterson by the vector *BC* and selecting the common patterns yields (*c*). Similarly, displacement by *AC* gives (*d*). Finally, superposition of (*c*) on (*d*) gives the original figure or its enantiomorph. This series of steps demonstrates that, in principle, complete structural information is contained in an SIR derivative.

(2.3.3.5) and (2.3.3.6) are indeed rather similar. The second term in (2.3.3.5) relates to the use of the search atoms in phasing and could be included in (2.3.3.6), provided the actual feedback sites in each of the n electron-density functions tested by $C_{\rm SIR}$ are omitted in turn. Thus, a systematic Patterson search and an SIR difference Fourier search are very similar in character and power.

2.3.3.6. Direct structure determination from difference Pattersons

The difference Patterson computed with coefficients $F_{HN}^2 - F_N^2$ contains information on the heavy atoms (*HH* vectors) and the macromolecular structure (*HL* vectors) (Section 2.3.3.3). If the scaling between the $|\mathbf{F}_{HN}|$ and $|\mathbf{F}_N|$ data sets is not perfect there will also be noise. Rossmann (1961*b*) was partially successful in determining the low-resolution horse haemoglobin structure by using a series of superpositions based on the known heavy-atom sites. Nevertheless, Patterson superposition methods have not been used for the structure determination of proteins owing to the successful error treatment of the isomorphous replacement method in reciprocal space. However, it is of some interest here for it gives an alternative insight into SIR phasing.

The deconvolution of an arbitrary molecule, represented as '?', from an $(F_{HN}^2 - F_N^2)$ Patterson, is demonstrated in Fig. 2.3.3.5. The original structure is shown in Fig. 2.3.3.5(*a*) and the corresponding Patterson in Fig. 2.3.3.5(*b*). Superposition with respect to one of the heavy-atom sites is shown in Fig. 2.3.3.5(*c*) and the other in Fig. 2.3.3.5(*d*). Both Figs. 2.3.3.5(*c*) and (*d*) contain a centre of symmetry because the use of only a single *HH* vector implies a centre of symmetry half way between the two sites. The centre is broken on combining information from all three sites (which together lack a centre of symmetry) by superimposing Figs. 2.3.3.5(*c*) and (*d*) to obtain either the original structure (Fig. 2.3.3.5*a*) or its enantiomorph. Thus it is clear, in principle, that there is sufficient information in a single isomorphous derivative data set, when used in conjunction with a native data set, to solve a structure completely. However, the procedure shown in Fig. 2.3.3.5 does not consider the accumulation of error in the selection of individual images when these intersect with another image. In this sense the reciprocal-space isomorphous replacement technique has greater elegance and provides more insight, whereas the alternative view given by the Patterson method was the original stimulus for the discovery of the SIR phasing technique (Blow & Rossmann, 1961).

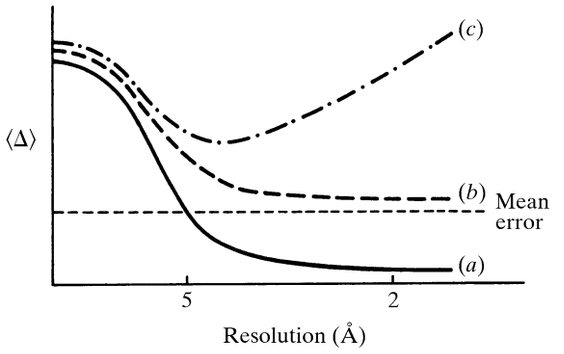


Fig. 2.3.3.6. A plot of mean isomorphous differences as a function of resolution. (*a*) The theoretical size of mean differences following roughly a Gaussian distribution. (*b*) The observed size of differences for a good isomorphous derivative where the smaller higher-order differences have been largely masked by the error of measurement. (*c*) Observed differences where 'lack of isomorphism' dominates beyond approximately 5 Å resolution.

Other Patterson functions for the deconvolution of SIR data have been proposed by Ramachandran & Raman (1959), as well as others. The principles are similar but the coefficients of the functions are optimized to emphasize various aspects of the signal representing the molecular structure.

2.3.3.7. Isomorphism and size of the heavy-atom substitution

It is insufficient to discuss Patterson techniques for locating heavy-atom substitutions without also considering errors of all kinds. First, it must be recognized that most heavy-atom labels are not a single atom but a small compound containing one or more heavy atoms. The compound itself will displace water or ions and locally alter the conformation of the protein or nucleic acid. Hence, a simple Gaussian approximation will suffice to represent individual heavy-atom scatterers responsible for the difference between native and heavy-atom derivatives. Furthermore, the heavy-atom compound often introduces small global structural changes which can be detected only at higher resolution. These problems were considered with some rigour by Crick & Magdoff (1956). In general, lack of isomorphism is exhibited by an increase in the size of the isomorphous differences with increasing resolution (Fig. 2.3.3.6).

Crick & Magdoff (1956) also derived the approximate expression

$$\sqrt{\frac{2N_{\rm H}}{N_{\rm P}}}\cdot\frac{f_{\rm H}}{f_{\rm P}}$$

to estimate the r.m.s. fractional change in intensity as a function of heavy-atom substitution. Here, $N_{\rm H}$ represents the number of heavy atoms attached to a protein (or other large molecule) which contains $N_{\rm P}$ light atoms. $f_{\rm H}$ and $f_{\rm P}$ are the scattering powers of the average heavy and protein atom, respectively. This function was tabulated by Eisenberg (1970) as a function of molecular weight (proportional to $N_{\rm P}$). For instance, for a single, fully substituted, Hg atom the formula predicts an r.m.s. intensity change of around 25% in a molecule of 100 000 Da. However, the error of measurement of a reflection intensity is likely to be arround 10% of I, implying perhaps an error of around 14% of I on a difference measurement. Thus, the isomorphous replacement difference measurement for almost half the reflections will be buried in error for this case.

Scaling of the different heavy-atom-derivative data sets onto a common relative scale is clearly important if error is to be reduced. Blundell & Johnson (1976, pp. 333–336) give a careful discussion of this subject. Suffice it to say here only that a linear scale factor is seldom acceptable as the heavy-atom-derivative crystals frequently suffer from greater disorder than the native crystals. The heavy-atom derivative should, in general, have a slightly larger mean value for the structure factors on account of the additional heavy atoms (Green *et al.*, 1954). The usual effect is to make $\sum|\mathbf{F}_{\rm NH}|^2/\sum|\mathbf{F}_{\rm N}|^2 \simeq 1.05$ (Phillips, 1966).

As the amount of heavy atom is usually unknown in a yet unsolved heavy-atom derivative, it is usual practice either to apply a scale factor of the form $k\exp[-B(\sin\theta/\lambda)^2]$ or, more generally, to use local scaling (Matthews & Czerwinski, 1975). The latter has the advantage of not making any assumption about the physical nature of the relative intensity decay with resolution.

2.3.4. Anomalous dispersion

2.3.4.1. Introduction

The physical basis for anomalous dispersion has been well reviewed by Ramaseshan & Abrahams (1975), James (1965), Cromer (1974) and Bijvoet (1954). As the wavelength of radiation approaches the absorption edge of a particular element, then an atom will disperse X-rays in a manner that can be defined by the complex scattering factor

$$f_0 + \Delta f' + i\Delta f'',$$

where f_0 is the scattering factor of the atom without the anomalous absorption and rescattering effect, $\Delta f'$ is the real correction term (usually negative), and $\Delta f''$ is the imaginary component. The real term $f_0 + \Delta f'$ is often written as f', so that the total scattering factor will be $f' + i\Delta f''$. Values of $\Delta f'$ and $\Delta f''$ are tabulated in *IT* IV (Cromer, 1974), although their precise values are dependent on the environment of the anomalous scatterer. Unlike f_0, $\Delta f'$ and $\Delta f''$ are almost independent of scattering angle as they are caused by absorption of energy in the innermost electron shells. Thus, the anomalous effect resembles scattering from a point atom.

The structure factor of index **h** can now be written as

$$\mathbf{F}_{\mathbf{h}} = \sum_{j=1}^{N} f_j' \exp(2\pi i\mathbf{h}\cdot\mathbf{x}_j) + i\sum_{j=1}^{N}\Delta f_j'' \exp(2\pi i\mathbf{h}\cdot\mathbf{x}_j). \qquad (2.3.4.1)$$

(Note that the only significant contributions to the second term are from those atoms that have a measurable anomalous effect at the chosen wavelength.)

Let us now write the first term as $A + iB$ and the second as $a + ib$. Then, from (2.3.4.1),

$$\mathbf{F} = (A + iB) + i(a + ib) = (A - b) + i(B + a). \qquad (2.3.4.2)$$

Therefore,

$$|\mathbf{F}_{\mathbf{h}}|^2 = (A - b)^2 + (B + a)^2$$

and similarly

$$|\mathbf{F}_{\bar{\mathbf{h}}}|^2 = (A + b)^2 + (-B + a)^2,$$

demonstrating that Friedel's law breaks down in the presence of anomalous dispersion. However, it is only for noncentrosymmetric reflections that $|\mathbf{F}_{\mathbf{h}}| \neq |\mathbf{F}_{\bar{\mathbf{h}}}|$.

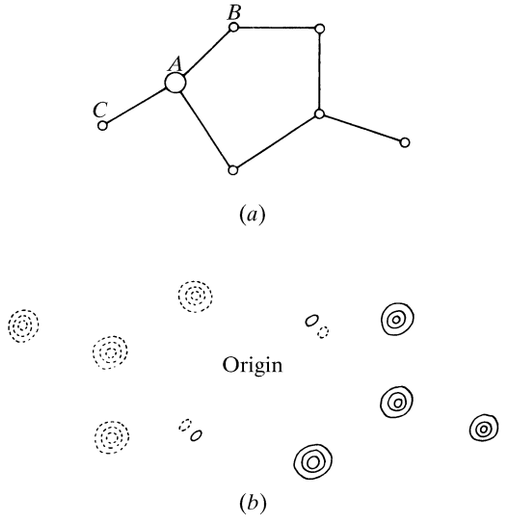


Fig. 2.3.4.1. (*a*) A model structure with an anomalous scatterer at *A*. (*b*) The corresponding $P_s(\mathbf{u})$ function showing positive peaks (full lines) and negative peaks (dashed lines). [Reprinted with permission from Woolfson (1970, p. 293).]

Now,

$$\rho(\mathbf{x}) = \frac{1}{V}\sum_{\mathbf{h}}^{\text{sphere}} \mathbf{F}_{\mathbf{h}} \exp(2\pi i \mathbf{h}\cdot\mathbf{x}).$$

Hence, by using (2.3.4.2) and simplifying,

$$\rho(\mathbf{x}) = \frac{2}{V}\sum_{\mathbf{h}}^{\text{hemisphere}} [(A\cos 2\pi\mathbf{h}\cdot\mathbf{x} - B\sin 2\pi\mathbf{h}\cdot\mathbf{x}) + i(a\cos 2\pi\mathbf{h}\cdot\mathbf{x} - b\sin 2\pi\mathbf{h}\cdot\mathbf{x})]. \quad (2.3.4.3)$$

The first term in (2.3.4.3) is the usual real Fourier expression for electron density, while the second term is an imaginary component due to the anomalous scattering of a few atoms in the cell.

2.3.4.2. The $P_s(\mathbf{u})$ function

Expression (2.3.4.3) gives the complex electron density expression in the presence of anomalous scatterers. A variety of Patterson-type functions can be derived from (2.3.4.3) for the determination of a structure. For instance, the $P_s(\mathbf{u})$ function gives vectors between the anomalous atoms and the 'normal' atoms.

From (2.3.4.1) it is easy to show that

$$\begin{aligned}\mathbf{F}_{\mathbf{h}}\mathbf{F}_{\mathbf{h}}^* &= |\mathbf{F}_{\mathbf{h}}|^2 \\ &= \sum_{i,j}(f_i' f_j' + f_i'' f_j'')\cos 2\pi\mathbf{h}\cdot(\mathbf{x}_i - \mathbf{x}_j) \\ &\quad + \sum_{i,j}(f_i' f_j'' - f_i'' f_j')\sin 2\pi\mathbf{h}\cdot(\mathbf{x}_i - \mathbf{x}_j).\end{aligned}$$

Therefore,

$$|\mathbf{F}_{\mathbf{h}}|^2 + |\mathbf{F}_{\bar{\mathbf{h}}}|^2 = 2\sum_{i,j}(f_i' f_j' + f_i'' f_j'')\cos 2\pi\mathbf{h}\cdot(\mathbf{x}_i - \mathbf{x}_j)$$

and

$$|\mathbf{F}_{\mathbf{h}}|^2 - |\mathbf{F}_{\bar{\mathbf{h}}}|^2 = 2\sum_{i,j}(f_i' f_j'' - f_i'' f_j')\sin 2\pi\mathbf{h}\cdot(\mathbf{x}_i - \mathbf{x}_j).$$

Let us now consider the significance of a Patterson in the presence of anomalous dispersion. The normal Patterson definition is given by

$$\begin{aligned}P(\mathbf{u}) &= \int_V \rho^*(\mathbf{x})\rho(\mathbf{x}+\mathbf{u})\,\mathrm{d}\mathbf{x} \\ &= \frac{1}{V^2}\sum_{\mathbf{h}}^{\text{sphere}} |\mathbf{F}_{\mathbf{h}}|^2 \exp(-2\pi i\mathbf{h}\cdot\mathbf{u}) \\ &\equiv P_c(\mathbf{u}) - iP_s(\mathbf{u}),\end{aligned}$$

where

$$P_c(\mathbf{u}) = \frac{2}{V}\sum_{\mathbf{h}}^{\text{hemisphere}} (|\mathbf{F}_{\mathbf{h}}|^2 + |\mathbf{F}_{\bar{\mathbf{h}}}|^2)\cos 2\pi\mathbf{h}\cdot\mathbf{u}$$

and

$$P_s(\mathbf{u}) = \frac{2}{V}\sum_{\mathbf{h}}^{\text{hemisphere}} (|\mathbf{F}_{\mathbf{h}}|^2 - |\mathbf{F}_{\bar{\mathbf{h}}}|^2)\sin 2\pi\mathbf{h}\cdot\mathbf{u}.$$

The $P_c(\mathbf{u})$ component is essentially the normal Patterson, in which the peak heights have been very slightly modified by the anomalous-scattering effect. That is, the peaks of $P_c(\mathbf{u})$ are proportional in height to $(f_i' f_j' + f_i'' f_j'')$.

The $P_s(\mathbf{u})$ component is more interesting. It represents vectors between the normal atoms in the unit cell and the anomalous scatterers, proportional in height to $(f_i' f_j'' - f_i'' f_j')$ (Okaya *et al.*, 1955). This function is antisymmetric with respect to the change of the direction of the diffraction vector. An illustration of the function is given in Fig. 2.3.4.1. In a unit cell containing N atoms, n of which are anomalous scatterers, the $P_s(\mathbf{u})$ function contains only $n(N-n)$ positive peaks and an equal number of negative peaks related to the former by anticentrosymmetry. The analysis of a $P_s(\mathbf{u})$ synthesis presents problems somewhat similar to those posed by a normal Patterson. The procedure has been used only rarely [*cf.* Moncrief & Lipscomb (1966) and Pepinsky *et al.* (1957)], probably because alternative procedures are available for small compounds and the solution of $P_s(\mathbf{u})$ is too complex for large biological molecules.

2.3.4.3. The position of anomalous scatterers

Anomalous scatterers can be used as an aid to phasing, when their positions are known. But the anomalous-dispersion differences (Bijvoet differences) can also be used to determine or confirm the heavy atoms which scatter anomalously (Rossmann, 1961*a*). Furthermore, the use of anomalous-dispersion information obviates the lack of isomorphism but, on the other hand, the differences are normally far smaller than those produced by a heavy-atom isomorphous replacement.

Consider a structure of many light atoms giving rise to the structure factor $\mathbf{F}_{\mathbf{h}}(N)$. In addition, it contains a few heavy atoms which have a significant anomalous-scattering effect. The non-anomalous component will be $\mathbf{F}_{\mathbf{h}}(H)$ and the anomalous component is $\mathbf{F}_{\mathbf{h}}''(H) = i(\Delta f''/f')\mathbf{F}_{\mathbf{h}}(H)$ (Fig. 2.3.4.2*a*). The total structure factor will be $\mathbf{F}_{\mathbf{h}}$. The Friedel opposite is constructed appropriately (Fig. 2.3.4.2*a*). Now reflect the Friedel opposite construction across the real axis of the Argand diagram (Fig.

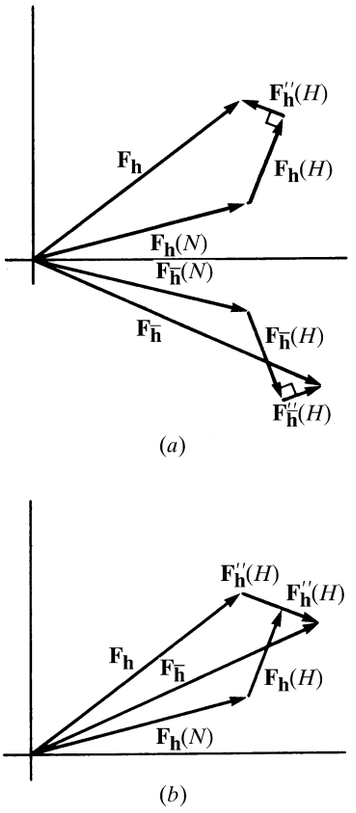


Fig. 2.3.4.2. Anomalous-dispersion effect for a molecule whose light atoms contribute $\mathbf{F}_{\mathbf{h}}(N)$ and heavy atom $\mathbf{F}_{\mathbf{h}}(H)$ with a small anomalous component of $\mathbf{F}_{\mathbf{h}}''(H)$, 90° ahead of the non-anomalous $\mathbf{F}_{\mathbf{h}}(H)$ component. In (*a*) is seen the construction for $\mathbf{F}_{\mathbf{h}}$ and $\mathbf{F}_{\bar{\mathbf{h}}}$. In (*b*) $\mathbf{F}_{\bar{\mathbf{h}}}$ has been reflected across the real axis.

2.3.4.2*b*). Let the difference in phase between $\mathbf{F}_{\mathbf{h}}$ and $\mathbf{F}_{\bar{\mathbf{h}}}$ be φ. Thus

$$4|\mathbf{F}_{\mathbf{h}}''(H)|^2 = |\mathbf{F}_{\mathbf{h}}|^2 + |\mathbf{F}_{\bar{\mathbf{h}}}|^2 - 2|\mathbf{F}_{\mathbf{h}}||\mathbf{F}_{\bar{\mathbf{h}}}|\cos\varphi,$$

but since φ is very small

$$|\mathbf{F}_{\mathbf{h}}''(H)|^2 \simeq \tfrac{1}{4}(|\mathbf{F}_{\mathbf{h}}| - |\mathbf{F}_{\bar{\mathbf{h}}}|)^2.$$

Hence, a Patterson with coefficients $(|\mathbf{F}_{\mathbf{h}}| - |\mathbf{F}_{\bar{\mathbf{h}}}|)^2$ will be equivalent to a Patterson with coefficients $|\mathbf{F}_{\mathbf{h}}''(H)|^2$ which is proportional to $|\mathbf{F}_{\mathbf{h}}(H)|^2$. Such a Patterson (Rossmann, 1961*a*) will have vectors between all anomalous scatterers with heights proportional to the number of anomalous electrons $\Delta f''$. This 'anomalous dispersion' Patterson function has been used to find anomalous scatterers such as iron (Smith *et al.*, 1983; Strahs & Kraut, 1968) and sulfur atoms (Hendrickson & Teeter, 1981). The anomalous signal from Se atoms in selenomethionine-substituted proteins has been found to be extremely powerful and is now routinely used for protein structure determinations (Hendrickson, 1991). Anomalous signals from halide ions or xenon atoms have also been used to solve protein structures (Dauter *et al.*, 2000; Nagem *et al.*, 2003; Schiltz *et al.*, 2003). The anomalous signal from sulfur atoms, though very small (Hendrickson & Teeter, 1981), has recently been applied successfully to solve several protein structures (Debreczeni *et al.*, 2003; Ramagopal *et al.*, 2003; Yang *et al.*, 2003).

It is then apparent that a Patterson with coefficients

$$\Delta F_{\mathrm{ANO}}^2 = (|\mathbf{F}_{\mathbf{h}}| - |\mathbf{F}_{\bar{\mathbf{h}}}|)^2$$

(Rossmann, 1961*a*), as well as a Patterson with coefficients

$$\Delta F_{\mathrm{ISO}}^2 = (|\mathbf{F}_{\mathrm{NH}}| - |\mathbf{F}_{\mathrm{H}}|)^2$$

(Rossmann, 1960; Blow, 1958), represent Pattersons of the heavy atoms. The $\Delta F_{\mathrm{ANO}}^2$ Patterson suffers from errors which may be larger than the size of the Bijvoet differences, while the $\Delta F_{\mathrm{ISO}}^2$ Patterson may suffer from partial lack of isomorphism. Hence, Kartha & Parthasarathy (1965) have suggested the use of the sum of these two Pattersons, which would then have coefficients $(\Delta F_{\mathrm{ANO}}^2 + \Delta F_{\mathrm{ISO}}^2)$.

However, given both SIR and anomalous-dispersion data, it is possible to make an accurate estimate of the $|\mathbf{F}_{\mathrm{H}}|^2$ magnitudes for use in a Patterson calculation [Blundell & Johnson (1976, p. 340), Matthews (1966), Singh & Ramaseshan (1966)]. In essence, the Harker phase diagram can be constructed out of three circles: the native amplitude and each of the two isomorphous Bijvoet differences, giving three circles in all (Blow & Rossmann, 1961) which should intersect at a single point thus resolving the ambiguity in the SIR data and the anomalous-dispersion data. Furthermore, the phase ambiguities are orthogonal; thus the two data sets are cooperative. It can be shown (Matthews, 1966; North, 1965) that

$$F_{\mathrm{N}}^2 = F_{\mathrm{NH}}^2 + F_{\mathrm{N}}^2 \mp \frac{2}{k}(16k^2F_{\mathrm{P}}^2F_{\mathrm{H}}^2 - \Delta I^2)^{1/2},$$

where $\Delta I = F_{\mathrm{NH}}^{+\,2} - F_{\mathrm{NH}}^{-\,2}$ and $k = \Delta f''/f'$. The sign in the third-term expression is $-$ when $|(\alpha_{\mathrm{NH}} - \alpha_{\mathrm{H}})| < \pi/2$ or $+$ otherwise. Since, in general, $|\mathbf{F}_{\mathrm{H}}|$ is small compared to $|\mathbf{F}_{\mathrm{N}}|$, it is reasonable to assume that the sign above is usually negative. Hence, the heavy-atom lower estimate (HLE) is usually written as

$$F_{\mathrm{HLE}}^2 = F_{\mathrm{NH}}^2 + F_{\mathrm{H}}^2 - \frac{2}{k}(16k^2F_{\mathrm{P}}^2F_{\mathrm{H}}^2 - \Delta I^2)^{1/2},$$

which is an expression frequently used to derive Patterson coefficients useful in the determination of heavy-atom positions when both SIR and anomalous-dispersion data are available.

2.3.4.4. Computer programs for automated location of atomic positions from Patterson maps

Several programs are currently used for automated systematic interpretation of (difference) Patterson maps to locate the positions of heavy atoms and/or anomalous scatterers from isomorphous replacement and anomalous-dispersion data (Weeks *et al.*, 2003). These include *Solve* (Terwilliger & Berendzen, 1999), *CNS* (Brünger *et al.*, 1998), *CCP4* (Collaborative Computational Project, Number 4, 1994) and *Patsol* (Tong & Rossmann, 1993). In these programs, sets of trial atomic positions (seeds) are produced based on one- and two-atom solutions to the Patterson map (see Section 2.3.2.5) (Grosse-Kunstleve & Brunger, 1999; Terwilliger *et al.*, 1987; Tong & Rossmann, 1993). Information from a translation search with a single atom can also be used in this process (Grosse-Kunstleve & Brunger, 1999). Scoring functions have been devised to identify the likely correct solutions, based on agreements with the Patterson map or the observed isomorphous or anomalous differences, as well as the quality of the resulting electron-density map (Terwilliger, 2003*b*; Terwilliger & Berendzen, 1999). The power of modern computers allows the rapid screening of a large collection of trial structures, and the correct solution is found automatically in many cases, even when there is a large number of atomic positions (Weeks *et al.*, 2003).

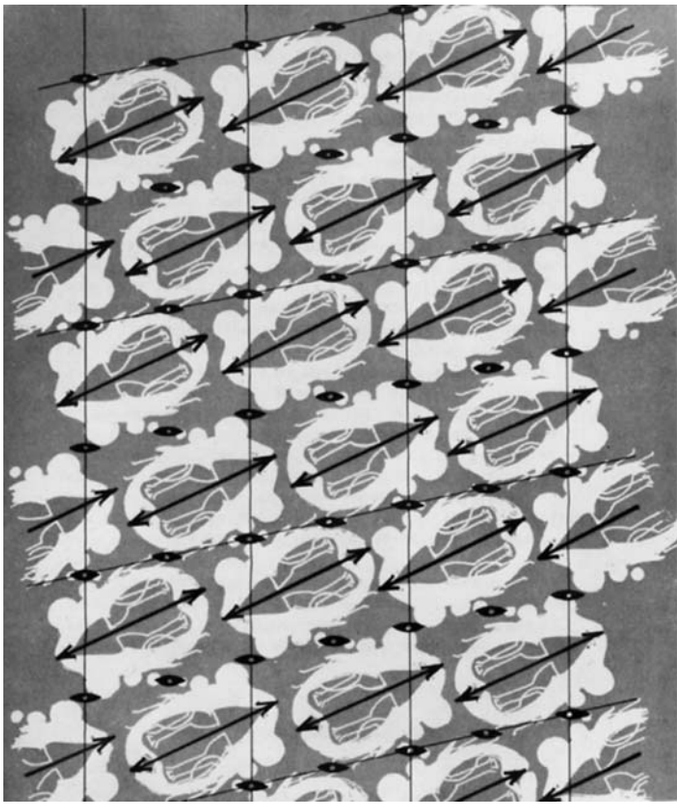


Fig. 2.3.5.1. The two-dimensional periodic design shows crystallographic twofold axes perpendicular to the page and local noncrystallographic rotation axes in the plane of the paper (design by Audrey Rossmann). [Reprinted with permission from Rossmann (1972, p. 8).]

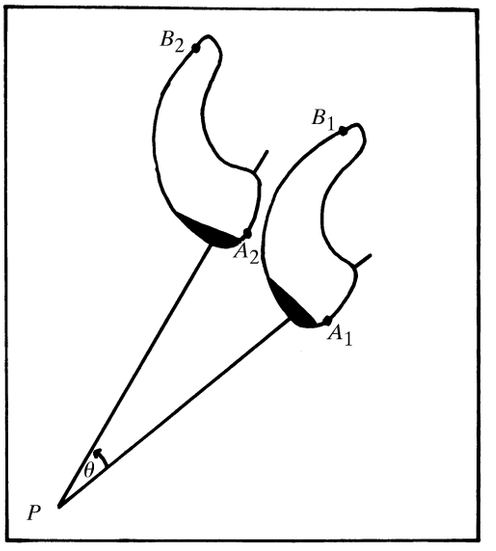


Fig. 2.3.5.2. The objects A_1B_1 and A_2B_2 are related by an improper rotation θ, since it is necessary to consider the sense of rotation to achieve superposition of the two objects. [Reprinted with permission from Rossmann (1972, p. 9).]

2.3.5. Noncrystallographic symmetry

2.3.5.1. Definitions

The interpretation of Pattersons can be helped by using various types of chemical or physical information. An obvious example is the knowledge that one or two heavy atoms per crystallographic asymmetric unit are present. Another example is the exploitation of a rigid chemical framework in a portion of a molecule (Nordman & Nakatsu, 1963; Burnett & Rossmann, 1971). One extremely useful constraint on the interpretation of Pattersons is noncrystallographic symmetry. Indeed, the structural solution of large biological assemblies such as viruses is only possible because of the natural occurrence of this phenomenon. The term 'molecular replacement' is used for methods that utilize noncrystallographic symmetry in the solution of structures [for earlier reviews see Rossmann (1972); Argos & Rossmann (1980); and Rossmann (1990, 2001)]. These methods, which are only partially dependent on Patterson concepts, are discussed in Sections 2.3.6–2.3.8.

Crystallographic symmetry applies to the whole of the three-dimensional crystal lattice. Hence, the symmetry must be expressed both in the lattice and in the repeating pattern within the lattice. In contrast, noncrystallographic symmetry is valid only within a limited volume about the noncrystallographic symmetry element. For instance, the noncrystallographic twofold axes in the plane of the paper of Fig. 2.3.5.1 are true only in the immediate vicinity of each local dyad. In contrast, the crystallographic twofold axes perpendicular to the plane of the paper (Fig. 2.3.5.1) apply to every point within the lattice. Two types of noncrystallographic symmetry can be recognized: proper and improper rotations. A proper symmetry element is independent of the sense of rotation, as, for example, a fivefold axis in an icosahedral virus; a rotation either left or right by one-fifth of a revolution will leave all parts of a given icosahedral shell (but not the whole crystal) in equivalent positions. Proper noncrystallographic symmetry can also be recognized by the existence of a closed point group within a defined volume of the lattice. Improper rotation axes are found when two molecules are arbitrarily oriented relative to each other in the same asymmetric unit or when they occur in two entirely different crystal lattices. For instance, in Fig. 2.3.5.2, the object A_1B_1 can be rotated by $+\theta$ about the axis at P to orient it identically with A_2B_2. However, the two objects will not be coincident after a rotation of A_1B_1 by $-\theta$ or of A_2B_2 by $+\theta$. The envelope around each noncrystallographic object must be known in order to define an improper rotation. In contrast, only the volume about the closed point group need be defined for proper noncrystallographic operations. Hence, the boundaries of the repeating unit need not correspond to chemically covalently linked units in the presence of proper rotations.

Translational components of noncrystallographic rotation elements are said to be 'precise' in a direction parallel to the axis and 'imprecise' perpendicular to the axis (Rossmann *et al.*, 1964). The position, but not direction, of a rotation axis is arbitrary. However, a convenient choice is one that leaves the translation perpendicular to the axis at zero after rotation (Fig. 2.3.5.3).

Noncrystallographic symmetry has been used as a tool in structural analysis primarily in the study of biological molecules. This is due to the propensity of proteins to form aggregates with closed point groups, as, for instance, viruses with 532 symmetry. At best, only part of such a point group can be incorporated into the crystal lattice. Since biological materials cannot contain inversion elements, all studies of noncrystallographic symmetries have been limited to rotational axes. Reflection planes and inversion centres could also be considered in the application of molecular replacement to nonbiological materials.

In this chapter, the relationship

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

will be used to describe noncrystallographic symmetry, where $\mathbf{x}$ and $\mathbf{x}'$ are position vectors, expressed as fractional coordinates, with respect to the crystallographic origin, $[\boldsymbol{C}]$ is a rotation matrix, and $\mathbf{d}$ is a translation vector. Crystallographic symmetry will be described as

$$\mathbf{x}' = [\boldsymbol{T}]\mathbf{x} + \mathbf{t},$$

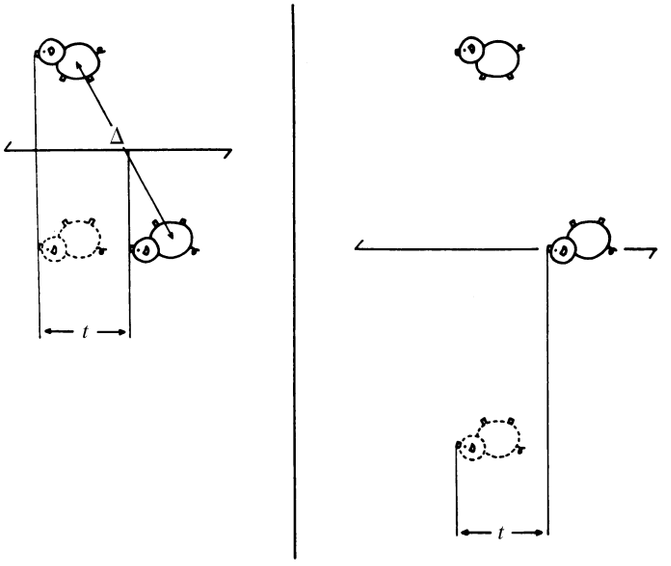


Fig. 2.3.5.3. The position of the twofold rotation axis which relates the two piglets is completely arbitrary. The diagram on the left shows the situation when the translation is parallel to the rotation axis. The diagram on the right has an additional component of translation perpendicular to the rotation axis, but the component parallel to the axis remains unchanged. [Reprinted from Rossmann *et al.* (1964).]

where $[\boldsymbol{T}]$ and $\mathbf{t}$ are the crystallographic rotation matrix and translation vector, respectively. The noncrystallographic asymmetric unit will be defined as having n copies within the crystallographic asymmetric unit, and the unit cell will be defined as having m crystallographic asymmetric units. Hence, there are $L = nm$ noncrystallographic asymmetric units within the unit cell. Clearly, the n noncrystallographic asymmetric units cannot completely fill the volume of one crystallographic asymmetric unit. The remaining space must be assumed to be empty or to be occupied by solvent molecules which disobey the noncrystallographic symmetry.

2.3.5.2. Interpretation of Pattersons in the presence of noncrystallographic symmetry

If noncrystallographic symmetry is present, an atom at a general position within the relevant volume will imply the presence of others within the same crystallographic asymmetric unit. If the noncrystallographic symmetry is known, then the positions of equivalent atoms may be generated from a single atomic position. The additional vector interactions which arise from crystallographically and noncrystallographically equivalent atoms in a crystal may be predicted and exploited in an interpretation of the Patterson function.

An object in real space which has a closed point group may incorporate some of its symmetry in the crystallographic symmetry. If there are l such objects in the cell, then there will be mn/l equivalent positions within each object. The 'self-vectors' formed between these positions within the object will be independent of the position of the objects. This distinction is important in that the self-vectors arising from atoms interacting with other atoms within a single particle may be correctly predicted without the knowledge of the particle centre position. In fact, this distinction may be exploited in a two-stage procedure in which an atom may be first located relative to the particle centre by use of the self-vectors and subsequently the particle may be positioned relative to crystallographic symmetry elements by use of the 'cross-vectors' (Table 2.3.5.1).

The interpretation of a heavy-atom difference Patterson for the holo-enzyme of lobster glyceraldehyde-3-phosphate dehydrogenase (GAPDH) provides an illustration of how the known noncrystallographic symmetry can aid the solution (Rossmann *et al.*, 1972; Buehner *et al.*, 1974). The GAPDH enzyme crystallized in a $P2_12_12_1$ cell (a = 149.0, b = 139.1, c = 80.7 Å) containing one tetramer per asymmetric unit. A rotation-function analysis had indicated the presence of three mutually perpendicular molecular twofold axes which suggested that the tetramer had 222 symmetry, and a locked rotation function determined the precise orientation of the tetramer relative to the crystal axes (see Table 2.3.5.2). Packing considerations led to assignment of a tentative particle centre near $\frac{1}{2}, \frac{1}{4}, Z$.

Table 2.3.5.1. *Possible types of vector searches*

Self-vectors	Cross-vectors	Dimension of search, n
(1) Locate single site relative to particle centre		$n = 3$
(2)	Use information from (1) to locate particle centre	$n \leq 3$
(3) Simultaneous search for both (1) and (2). In general this is a six-dimensional search but may be simplified when particle is on a crystallographic symmetry axis		$3 \leq n \leq 6$
(4) Given (1) for more than one site, find all vectors within particle		$n = 3$
(5) Given information from (3), locate additional site using complete vector set		$n = 3$

An isomorphous difference Patterson was calculated for the K_2HgI_4 derivative of GAPDH using data to a resolution of 6.8 Å. From an analysis of the three Harker sections, a tentative first heavy-atom position was assigned (atom A_2 at x, y, z). At this juncture, the known noncrystallographic symmetry was used to obtain a full interpretation. From Table 2.3.5.2 we see that molecular axis 2 will generate a second heavy atom with coordinates roughly $\frac{1}{4} + y, -\frac{1}{4} + x, 2Z - z$ (if the molecular centre was assumed to be at $\frac{1}{2}, \frac{1}{4}, Z$). Starting from the tentative coordinates of site A_2, the site A_1 related by molecular axis 1 was detected at about the predicted position and the second site A_1 generated acceptable cross-vectors with the earlier determined site A_2. Further examination enabled the completion of the set of four noncrystallographically related heavy-atom sites, such that all predicted Patterson vectors were acceptable and all four sites placed the molecular centre in the same position. Following refinement of these four sites, the corresponding SIR phases were used to find an additional set of four sites in this compound as well as in a number of other derivatives. The multiple isomorphous replacement phases, in conjunction with real-space electron-density averaging of the noncrystallographically related units, were then sufficient to solve the GAPDH structure.

When investigators studied larger macromolecular aggregates such as the icosahedral viruses, which have 532 point symmetry, systematic methods were developed for utilizing the noncrystallographic symmetry to aid in locating heavy-atom sites in isomorphous heavy-atom derivatives. Argos & Rossmann (1974, 1976) introduced an exhaustive Patterson search procedure for a single heavy-atom site within the noncrystallographic asymmetric unit which has been successfully applied to the interpretation of both virus [satellite tobacco necrosis virus (STNV) (Lentz *et al.*, 1976), southern bean mosaic virus (Rayment *et al.*, 1978), alfalfa mosaic virus (Fukuyama *et al.*, 1983), cowpea mosaic virus

Table 2.3.5.2. *Orientation of the glyceraldehyde-3-phosphate dehydrogenase molecular twofold axis in the orthorhombic cell*

Rotation axes	Polar coordinates (°)		Cartesian coordinates (direction cosines)		
	ψ	φ	u	v	w
1	45.0	−7.0	0.7018	0.7071	−0.0862
2	180.0–55.0	38.6	0.6402	−0.5736	0.5111
3	180.0–66.0	−70.6	0.3035	−0.4067	−0.8616

Table 2.3.6.1. *Different types of uses for the rotation function*

Type of rotation function	Pattersons to be compared		Purpose
	P_1	P_2	
Self	Unknown structure	Unknown structure, same cell	Finds orientation of noncrystallographic axes
Cross	Unknown structure	Unknown structure in different cell	Finds relative orientation of unknown molecules
Cross	Unknown structure	Known structure in large cell to avoid overlap of self-Patterson vectors	Determines orientation of unknown structure as preliminary to positioning and subsequent phasing with known molecule

(Stauffacher *et al.*, 1987)] and enzyme [catalase (Murthy *et al.*, 1981)] heavy-atom difference Pattersons. This procedure has also been implemented in the program *Patsol* (Tong & Rossmann, 1993). A heavy atom is placed in turn at all plausible positions within the volume of the noncrystallographic asymmetric unit and the corresponding vector set is constructed from the resulting constellation of heavy atoms. Argos & Rossmann (1976) found a spherical polar coordinate search grid to be convenient for spherical viruses. After all vectors for the current search position are predicted, the vectors are allocated to the nearest grid point and the list is sorted to eliminate recurring ones. The criterion used by Argos & Rossmann for selecting a solution is that the sum

$$S = \sum_{i=1}^{N} P_i - NP_{\rm av}$$

of the lookup Patterson density values P_i achieves a high value for a correct heavy-atom position. The sum is corrected for the carpet of cross-vectors by the second term in the sum.

An additional criterion, which has been found useful for discriminating correct solutions, is a unit vector density criterion (Arnold *et al.*, 1987)

$$U = \sum_{i=1}^{N} (P_i/n_i) \Big/ N,$$

where n_i is the number of vectors expected to contribute to the Patterson density value P_i (Arnold *et al.*, 1987). This criterion can be especially valuable for detecting correct solutions at special search positions, such as an icosahedral fivefold axis, where the number of vector lookup positions may be drastically reduced owing to the higher symmetry. An alternative, but equivalent, method for locating heavy-atom positions from isomorphous difference data is discussed in Section 2.3.3.5.

Even for a single heavy-atom site at a general position in the simplest icosahedral or ($T = 1$) virus, there are 60 equivalent heavy atoms in one virus particle. The number of unique vectors corresponding to this self-particle vector set will depend on the crystal symmetry but may be as many as $(60)(59)/2 = 1770$ for a virus particle at a general crystallographic position. Such was the case for the STNV crystals which were in space group $C2$ containing four virus particles at general positions. The method of Argos & Rossmann was applied successfully to a solution of the K_2HgI_4 derivative of STNV using a 10 Å resolution difference Patterson. Application of the noncrystallographic symmetry vector search procedure to a $K_2Au(CN)_2$ derivative of human rhinovirus 14 (HRV14) crystals (space group $P2_13$, $Z = 4$) has succeeded in establishing both the relative positions of heavy atoms within one particle and the positions of the virus particles relative to the crystal symmetry elements (Arnold *et al.*, 1987). The particle position was established by incorporating inter-particle vectors in the search and varying the particle position along the crystallographic threefold axis until the best fit for the predicted vector set was achieved.

Conversely, the knowledge that the heavy-atom positions, especially the Se atoms in a selenomethionyl protein, should obey the noncrystallographic symmetry can be used to deduce the nature, orientation and position of the noncrystallographic symmetry in the crystal unit cell, with either manual or automated procedures (Buehner *et al.*, 1974; Lu, 1999; Terwilliger, 2002*a*). The noncrystallographic symmetry can also serve as a powerful tool for refining the phase information derived from the heavy-atom positions (Buehner *et al.*, 1974).

2.3.6. Rotation functions

2.3.6.1. Introduction

The rotation function is designed to detect noncrystallographic rotational symmetry (see Table 2.3.6.1). The normal rotation function definition is given as (Rossmann & Blow, 1962)

$$R = \int_U P_1(\mathbf{u}) \cdot P_2(\mathbf{u}')\, {\rm d}\mathbf{u}, \qquad (2.3.6.1)$$

where P_1 and P_2 are two Pattersons and U is an envelope centred at the superimposed origins. This convolution therefore measures the degree of similarity, or 'overlap', between the two Pattersons when P_2 has been rotated relative to P_1 by an amount defined by

$$\mathbf{u}' = [\boldsymbol{C}]\mathbf{u}. \qquad (2.3.6.2)$$

The elements of $[\boldsymbol{C}]$ will depend on three rotation angles $(\theta_1, \theta_2, \theta_3)$. Thus, R is a function of these three angles. Alternatively, the matrix $[\boldsymbol{C}]$ could be used to express mirror symmetry, permitting searches for noncrystallographic mirror or glide planes.

The basic concepts were first clearly stated by Rossmann & Blow (1962), although intuitive uses of the rotation function had been considered earlier. Hoppe (1957*b*) had also hinted at a convolution of the type given by (2.3.6.1) to find the orientation of known molecular fragments and these ideas were implemented by Huber (1965).

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 away 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 centred at the origin and limited 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. Similarly,

the superposition of the molecular self-Patterson derived from different crystal forms can provide the relative orientation of the two crystals when the molecules are aligned.

While it would be possible to evaluate R by interpolating in P_2 and forming the point-by-point product with P_1 within the volume U for every combination of θ_1, θ_2 and θ_3, such a process is tedious and requires large computer storage for the Pattersons. Instead, the process is usually performed in reciprocal space where the number of independent structure amplitudes which form the Pattersons is about one-thirtieth of the number of Patterson grid points. Thus, the computation of a rotation function is carried out directly on the structure amplitudes, while the overlap definition (2.3.6.1) simply serves as a physical basis for the technique.

The derivation of the reciprocal-space expression depends on the expansion of each Patterson either as a Fourier summation, the conventional approach of Rossmann & Blow (1962), or as a sum of spherical harmonics in Crowther's (1972) analysis. The conventional and mathematically easier treatment is discussed presently, but the reader is referred also to Section 2.3.6.5 for Crowther's elegant approach. The latter leads to a rapid technique for performing the computations, about one hundred times faster than conventional methods.

Let, omitting constant coefficients,

$$P_1(\mathbf{u}) = \sum_{\mathbf{h}} |\mathbf{F}_{\mathbf{h}}|^2 \exp(2\pi i \mathbf{h} \cdot \mathbf{u})$$

and

$$P_2(\mathbf{u}') = \sum_{\mathbf{p}} |\mathbf{F}_{\mathbf{p}}|^2 \exp(2\pi i \mathbf{p} \cdot \mathbf{u}').$$

From (2.3.6.2) it follows that

$$P_2(\mathbf{u}') = \sum_{\mathbf{p}} |\mathbf{F}_{\mathbf{p}}|^2 \exp(2\pi i \mathbf{p}[\boldsymbol{C}] \cdot \mathbf{u}),$$

and, hence, by substitution in (2.3.6.1)

$$\begin{aligned} R(\theta_1, \theta_2, \theta_3) = \int_U & \left[\sum_{\mathbf{h}} |\mathbf{F}_{\mathbf{h}}|^2 \exp(2\pi i \mathbf{h} \cdot \mathbf{u}) \right] \\ & \times \left[\sum_{\mathbf{p}} |\mathbf{F}_{\mathbf{p}}|^2 \exp(2\pi i \mathbf{p}[\boldsymbol{C}] \cdot \mathbf{u}) \right] \mathrm{d}\mathbf{u} \\ = U \sum_{\mathbf{h}} |\mathbf{F}_{\mathbf{h}}|^2 & \left(\sum_{\mathbf{p}} |\mathbf{F}_{\mathbf{p}}|^2 G_{\mathbf{hp}} \right), \end{aligned} \tag{2.3.6.3}$$

where

$$U G_{\mathbf{hp}} = \int_U \exp\{2\pi i (\mathbf{h} + \mathbf{p}[\boldsymbol{C}]) \cdot \mathbf{u}\} \, \mathrm{d}\mathbf{u}.$$

When the volume U is a sphere, $G_{\mathbf{hp}}$ has the analytical form

$$G_{\mathbf{hp}} = \frac{3(\sin\theta - \theta\cos\theta)}{\theta^3}, \tag{2.3.6.4}$$

where $\theta = 2\pi HR$ and $H = \mathbf{h} + \mathbf{p}[\boldsymbol{C}]$. G is a spherical interference function whose form is shown in Fig. 2.3.6.1.

The expression (2.3.6.3) represents the rotation function in reciprocal space. If $\mathbf{h}' = [\boldsymbol{C}^T]\mathbf{p}$ in the argument of $G_{\mathbf{hp}}$, then $\mathbf{h}'$ can be seen as the point in reciprocal space to which $\mathbf{p}$ is rotated by

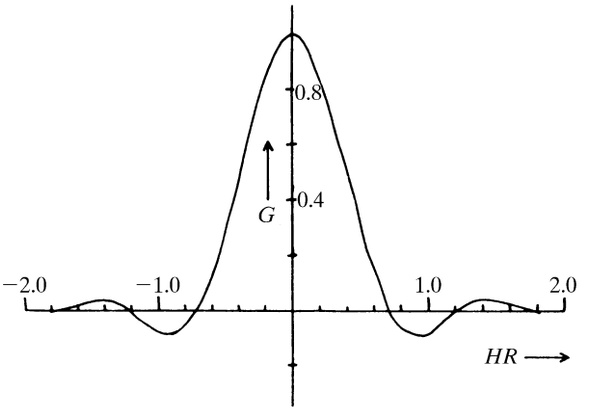


Fig. 2.3.6.1. Shape of the interference function G for a spherical envelope of radius R at a distance H from the reciprocal-space origin. [Reprinted from Rossmann & Blow (1962).]

$[\boldsymbol{C}]$. Only for those integral reciprocal-lattice points which are close to $-\mathbf{h}'$ will $G_{\mathbf{hp}}$ be of an appreciable size (Fig. 2.3.6.1). Thus, the number of significant terms is greatly reduced in the summation over $\mathbf{p}$ for every value of $\mathbf{h}$, making the computation of the rotation function manageable.

The radius of integration R should be approximately equal to or a little smaller than the molecular diameter. If R were roughly equal to the length of a lattice translation, then the separation of reciprocal-lattice points would be about $1/R$. Hence, when H is equal to one reciprocal-lattice separation, $HR \simeq 1$, and G is thus quite small. Indeed, all terms with $HR > 1$ might well be neglected. Thus, in general, the only terms that need be considered are those where $-\mathbf{h}'$ is within one lattice point of $\mathbf{h}$. However, in dealing with a small molecular fragment for which R is small compared to the unit-cell dimensions, more reciprocal-lattice points must be included for the summation over $\mathbf{p}$ in the rotation-function expression (2.3.6.3).

In practice, the equation

$$\mathbf{h} + \mathbf{h}' = 0,$$

that is

$$[\boldsymbol{C}^T]\mathbf{p} = -\mathbf{h}$$

or

$$\mathbf{p} = [\boldsymbol{C}^T]^{-1}(-\mathbf{h}), \tag{2.3.6.5}$$

determines $\mathbf{p}$, given a set of Miller indices $\mathbf{h}$. This will give a non-integral set of Miller indices. The terms included in the inner summation of (2.3.6.3) will be integral values of $\mathbf{p}$ around the non-integral lattice point found by solving (2.3.6.5).

Details of the conventional program were given by Tollin & Rossmann (1966) and follow the principles outlined above. They discussed various strategies as to which crystal should be used to calculate the first ($\mathbf{h}$) and second ($\mathbf{p}$) Patterson. Rossmann & Blow (1962) noted that the factor $\sum_{\mathbf{p}} |\mathbf{F}_{\mathbf{p}}|^2 G_{\mathbf{hp}}$ in expression (2.3.6.3) represents an interpolation of the squared transform of the self-Patterson of the second ($\mathbf{p}$) crystal. Thus, the rotation function is a sum of the products of the two molecular transforms taken over all the $\mathbf{h}$ reciprocal-lattice points. Lattman & Love (1970) therefore computed the molecular transform explicitly and stored it in the computer, sampling it as required by the rotation operation. A discussion on the suitable choice of variables in the computation of rotation functions has been given by Lifchitz (1983).

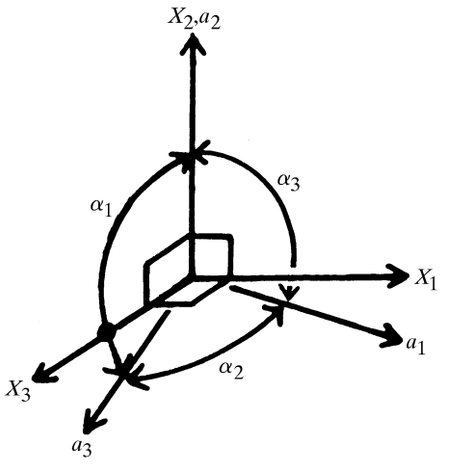


Fig. 2.3.6.2. Relationships of the orthogonal axes X_1, X_2, X_3 to the crystallographic axes a_1, a_2, a_3. [Reprinted from Rossmann & Blow (1962).]

2.3.6.2. Matrix algebra

The initial step in the rotation-function procedure involves the orthogonalization of both crystal systems. Thus, if fractional coordinates in the first crystal system are represented by **x**, these can be orthogonalized by a matrix $[\boldsymbol{\beta}]$ to give the coordinates **X** in units of length (Fig. 2.3.6.2); that is,

$$\mathbf{X} = [\boldsymbol{\beta}]\mathbf{x}.$$

If the point **X** is rotated to the point $\mathbf{X}'$, then

$$\mathbf{X}' = [\boldsymbol{\rho}]\mathbf{X}, \qquad (2.3.6.6)$$

where $\boldsymbol{\rho}$ represents the rotation matrix relating the two vectors in the orthogonal system. Finally, $\mathbf{X}'$ is converted back to fractional coordinates measured along the oblique cell dimension in the second crystal by

$$\mathbf{x}' = [\boldsymbol{\alpha}]\mathbf{X}'.$$

Thus, by substitution,

$$\mathbf{x}' = [\boldsymbol{\alpha}][\boldsymbol{\rho}]\mathbf{X} = [\boldsymbol{\alpha}][\boldsymbol{\rho}][\boldsymbol{\beta}]\mathbf{x}, \qquad (2.3.6.7)$$

and by comparison with (2.3.6.2) it follows that

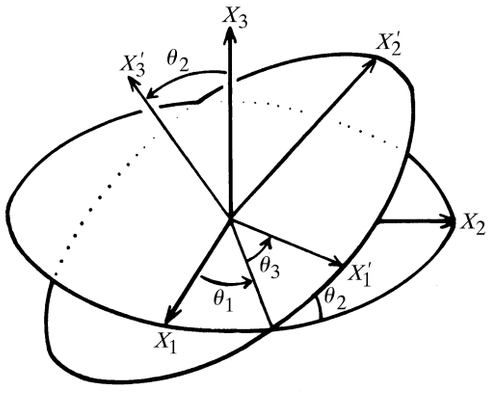


Fig. 2.3.6.3. Eulerian angles $\theta_1, \theta_2, \theta_3$ relating the rotated axes X_1', X_2', X_3' to the original unrotated orthogonal axes X_1, X_2, X_3. [Reprinted from Rossmann & Blow (1962).]

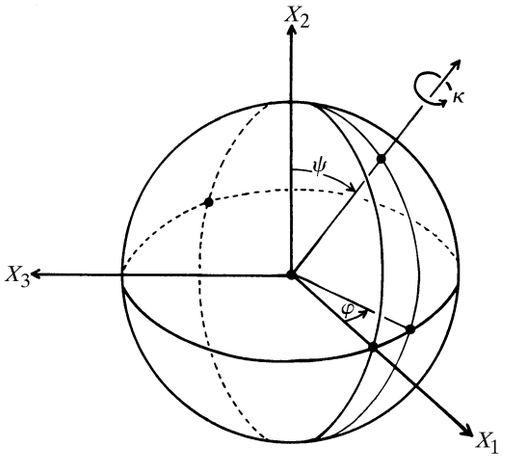


Fig. 2.3.6.4. Variables ψ and φ are polar coordinates which specify a direction about which the axes may be rotated through an angle κ. [Reprinted from Rossmann & Blow (1962).]

$$[\boldsymbol{C}] = [\boldsymbol{\alpha}][\boldsymbol{\rho}][\boldsymbol{\beta}].$$

Fig. 2.3.6.2 shows the mode of orthogonalization used by Rossmann & Blow (1962). With their definition it can be shown that

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

and

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

where $\cos\omega = (\cos\alpha_2 - \cos\alpha_1 \cos\alpha_3)/(\sin\alpha_1 \sin\alpha_3)$ with $0 \le \omega < \pi$. For a Patterson compared with itself, $[\boldsymbol{\alpha}] = [\boldsymbol{\beta}]^{-1}$.

An alternative mode of orthogonalization, used by the Protein Data Bank and most programs, is to align the a_1 axis of the unit cell with the Cartesian X_1 axis, and to align the a_3^* axis with the Cartesian X_3 axis. With this definition, the orthogonalization matrix is

$$[\boldsymbol{\beta}] = \begin{pmatrix} a_1 & a_2 \cos\alpha_3 & a_3 \cos\alpha_2 \\ 0 & a_2 \sin\alpha_3 & -a_3 \sin\alpha_2 \cos\alpha_1^* \\ 0 & 0 & a_3 \sin\alpha_2 \sin\alpha_1^* \end{pmatrix}.$$

Other modes of orthogonalization are also possible, some of which are supported in the program *GLRF* (Tong & Rossmann, 1990, 1997).

Both spherical (κ, ψ, φ) and Eulerian $(\theta_1, \theta_2, \theta_3)$ angles are used in evaluating the rotation function. The usual definitions employed are given diagrammatically in Figs. 2.3.6.3 and 2.3.6.4. They give rise to the following rotation matrices.

(*a*) Matrix $[\boldsymbol{\rho}]$ in terms of Eulerian angles $\theta_1, \theta_2, \theta_3$:

$$\begin{pmatrix} -\sin\theta_1\cos\theta_2\sin\theta_3 + \cos\theta_1\cos\theta_3 & \cos\theta_1\cos\theta_2\sin\theta_3 + \sin\theta_1\cos\theta_3 & \sin\theta_2\sin\theta_3 \\ -\sin\theta_1\cos\theta_2\cos\theta_3 - \cos\theta_1\sin\theta_3 & \cos\theta_1\cos\theta_2\cos\theta_3 - \sin\theta_1\sin\theta_3 & \sin\theta_2\cos\theta_3 \\ \sin\theta_1\sin\theta_2 & -\cos\theta_1\sin\theta_2 & \cos\theta_2 \end{pmatrix}$$

and (b) matrix $[\boldsymbol{\rho}]$ in terms of rotation angle κ and the spherical polar coordinates ψ, φ:

$$\begin{pmatrix} \cos\kappa + \sin^2\psi\cos^2\varphi(1-\cos\kappa) & \sin\psi\cos\psi\cos\varphi(1-\cos\kappa) + \sin\psi\sin\varphi\sin\kappa & -\sin^2\psi\cos\varphi\sin\varphi(1-\cos\kappa) + \cos\psi\sin\kappa \\ \sin\psi\cos\psi\cos\varphi(1-\cos\kappa) - \sin\psi\sin\varphi\sin\kappa & \cos\kappa + \cos^2\psi(1-\cos\kappa) & -\sin\psi\cos\psi\sin\varphi(1-\cos\kappa) - \sin\psi\cos\varphi\sin\kappa \\ -\sin^2\psi\sin\varphi\cos\varphi(1-\cos\kappa) - \cos\psi\sin\kappa & -\sin\psi\cos\psi\sin\varphi(1-\cos\kappa) + \sin\psi\cos\varphi\sin\kappa & \cos\kappa + \sin^2\psi\sin^2\varphi(1-\cos\kappa) \end{pmatrix}$$

Alternatively, (b) can be expressed as

$$\begin{pmatrix} \cos\kappa + u^2(1-\cos\kappa) & uv(1-\cos\kappa) - w\sin\kappa & uw(1-\cos\kappa) + v\sin\kappa \\ vu(1-\cos\kappa) + w\sin\kappa & \cos\kappa + v^2(1-\cos\kappa) & uw(1-\cos\kappa) - u\sin\kappa \\ wu(1-\cos\kappa) - v\sin\kappa & wv(1-\cos\kappa) + u\sin\kappa & \cos\kappa + w^2(1-\cos\kappa) \end{pmatrix},$$

where u, v and w are the direction cosines of the rotation axis given by

$$\begin{aligned} u &= \sin\psi\cos\varphi, \\ v &= \cos\psi, \\ w &= -\sin\psi\sin\varphi. \end{aligned}$$

This latter form also demonstrates that the trace of a rotation matrix is $2\cos\kappa + 1$.

The relationship between the two sets of variables established by comparison of the elements of the two matrices yields

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

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 $(\theta_1, \theta_2, \theta_3)$.

Another definition for the polar angles is also commonly used. In this definition, the angle ψ is measured from the Cartesian Z (X_3) axis, instead of the Y (X_2) axis. As most space groups have the unique axis along a_3, the ψ angle will define the inclination relative to the unique axis of the space group with this definition.

2.3.6.3. Symmetry

In analogy with crystal lattices, the rotation function is periodic and contains symmetry. The rotation function has a cell whose periodicity is 2π in each of its three angles. This may be written as

$$R(\theta_1, \theta_2, \theta_3) \equiv R(\theta_1 + 2\pi n_1, \theta_2 + 2\pi n_2, \theta_3 + 2\pi n_3)$$

or

$$R(\kappa, \psi, \varphi) \equiv R(\kappa + 2\pi n_1, \psi + 2\pi n_2, \varphi + 2\pi n_3),$$

where n_1, n_2 and n_3 are integers. A redundancy in the definition of either set of angles leads to the equivalence of the following points:

$$R(\theta_1, \theta_2, \theta_3) \equiv R(\theta_1 + \pi, -\theta_2, \theta_3 + \pi) \quad \text{in Eulerian space}$$

or

$$R(\kappa, \psi, \varphi) \equiv R(\kappa, 2\pi - \psi, \varphi + \pi) \quad \text{in polar space.}$$

These relationships imply an n glide plane perpendicular to θ_2 for Eulerian space or a φ glide plane perpendicular to ψ in polar space.

In addition, the Laue symmetry of the two Pattersons themselves must be considered. This problem was first discussed by Rossmann & Blow (1962) and later systematized by Tollin *et al.* (1966), Burdina (1970, 1971, 1973) and Rao *et al.* (1980). A closely related problem was considered by Hirshfeld (1968). The rotation function will have the same value whether the Patterson density at $\mathbf{X}$ or $[\boldsymbol{T}_i]\mathbf{X}$ in the first crystal is multiplied by the Patterson density at $\mathbf{X}'$ or $[\boldsymbol{T}_j]\mathbf{X}'$ in the second crystal. $[\boldsymbol{T}_i]$ and $[\boldsymbol{T}_j]$ refer to the ith and jth crystallographic rotations in the orthogonalized coordinate systems of the first and second crystal, respectively. Hence, from (2.3.6.6)

$$([\boldsymbol{T}_j]\mathbf{X}') = [\boldsymbol{\rho}]([\boldsymbol{T}_i]\mathbf{X})$$

or

$$\mathbf{X}' = [\boldsymbol{T}_j^T][\boldsymbol{\rho}][\boldsymbol{T}_i]\mathbf{X}.$$

Thus, it is necessary to find angular relationships which satisfy the relation

$$[\boldsymbol{\rho}] = [\boldsymbol{T}_j^T][\boldsymbol{\rho}][\boldsymbol{T}_i]$$

for given Patterson symmetries. Tollin *et al.* (1966) show that the Eulerian angular equivalences can be expressed in terms of the Laue symmetries of each Patterson (Table 2.3.6.2).

The example given by Tollin *et al.* (1966) is instructive in the use of Table 2.3.6.2. They consider the determination of the Eulerian space group when P_1 has symmetry *Pmmm* and P_2 has symmetry $P2/m$. These Pattersons contain the proper rotation groups 222 and 2 (parallel to **b**), respectively. Inspection of Table 2.3.6.2 shows that these symmetries produce the following Eulerian relationships:

Table 2.3.6.2. *Eulerian symmetry elements for all possible types of space-group rotations*

Axis	Direction	First crystal	Second crystal
1		$(\pi+\theta_1, -\theta_2, \pi+\theta_3)$	$(\pi+\theta_1, -\theta_2, \pi+\theta_3)$
2	[010]	$(\pi-\theta_1, \pi+\theta_2, \theta_3)$	$(\theta_1, \pi+\theta_2, \pi-\theta_3)$
2	[001]	$(\pi+\theta_1, \theta_2, \theta_3)$	$(\theta_1, \theta_2, \pi+\theta_3)$
4	[001]	$(-\pi/2+\theta_1, \theta_2, \theta_3)$	$(\theta_1, \theta_2, \pi/2+\theta_3)$
3	[001]	$(-2\pi/3+\theta_1, \theta_2, \theta_3)$	$(\theta_1, \theta_2, 2\pi/3+\theta_3)$
6	[001]	$(-\pi/3+\theta_1, \theta_2, \theta_3)$	$(\theta_1, \theta_2, \pi/3+\theta_3)$
2†	[110]	$(3\pi/2-\theta_1, \pi-\theta_2, \pi+\theta_3)$	$(\pi+\theta_1, \pi-\theta_2, -3\pi/2-\theta_3)$

† This axis is not unique (that is, it can always be generated by two other unique axes), but is included for completeness.

Table 2.3.6.3. *Numbering of the rotation-function space groups*

The Laue group of the rotated Patterson map P_1 is chosen from the left column and the Laue group of P_2 is chosen from the upper row.

	1	2/*m*, *b* axis unique	2/*m*, *c* axis unique	*mmm*	4/*m*	4/*mmm*	3	3*m*	6/*m*	6/*mmm*
1	1	11	21	31	41	51	61	71	81	91
2/*m*, *b* axis unique	2	12	22	32	42	52	62	72	82	92
2/*m*, *c* axis unique	3	13	23	33	43	53	63	73	83	93
mmm	4	14	24	34	44	54	64	74	84	94
4/*m*	5	15	25	35	45	55	65	75	85	95
4/*mmm*	6	16	26	36	46	56	66	76	86	96
3	7	17	27	37	47	57	67	77	87	97
3*m*	8	18	28	38	48	58	68	78	88	98
6/*m*	9	19	29	39	49	59	69	79	89	99
6/*mmm*	10	20	30	40	50	60	70	80	90	100

(*a*) In the first crystal (*Pmmm*):

$\theta_1\theta_2\theta_3 \rightarrow \pi+\theta_1, -\theta_2, \pi+\theta_3$ (onefold axis)
$\theta_1\theta_2\theta_3 \rightarrow \pi-\theta_1, \pi+\theta_2, \theta_3$ (twofold axis parallel to **b**)
$\theta_1\theta_2\theta_3 \rightarrow \pi+\theta_1, \theta_2, \theta_3$ (twofold axis parallel to **c**).

(*b*) In the second crystal (*P*2/*m*):

$\theta_1\theta_2\theta_3 \rightarrow \pi+\theta_1, -\theta_2, \pi+\theta_3$ (onefold axis)
$\theta_1\theta_2\theta_3 \rightarrow \theta_1, \pi+\theta_2, \pi-\theta_3$ (twofold axis parallel to **b**).

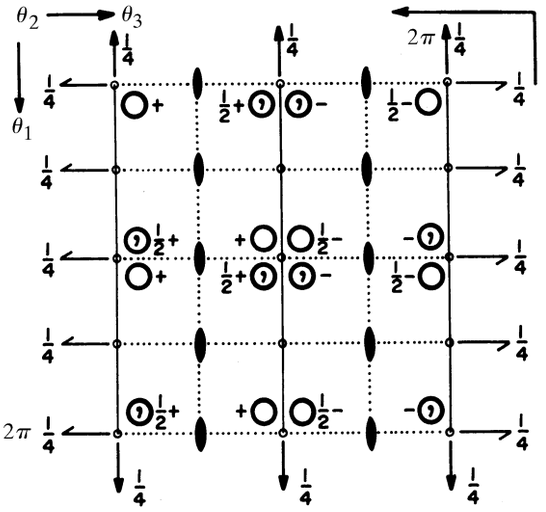


Fig. 2.3.6.5. Rotation space group diagram for the rotation function of a *Pmmm* Patterson function (P_1) against a *P*2/*m* Patterson function (P_2). The Eulerian angles $\theta_1, \theta_2, \theta_3$ repeat themselves after an interval of 2π. Heights above the plane are given in fractions of a revolution. [Reprinted from Tollin *et al.* (1966).]

When these symmetry operators are combined two cells result, each of which has the space group *Pbcb* (Fig. 2.3.6.5). The asymmetric unit within which the rotation function need be evaluated is found from a knowledge of the Eulerian space group. In the above example, the limits of the asymmetric unit are $0 \le \theta_1 \le \pi/2$, $0 \le \theta_2 \le \pi$ and $0 \le \theta_3 \le \pi/2$.

Nonlinear transformations occur when using Eulerian symmetries for threefold axes along [111] (as in the cubic system) or when using polar coordinates. Hence, Eulerian angles are far more suitable for a derivation of the limits of the rotation-function asymmetric unit. However, when searching for given molecular axes, where some plane of κ need be explored, polar angles are more useful.

Rao *et al.* (1980) have determined all possible rotation-function Eulerian space groups, except for combinations with Pattersons of cubic space groups. They numbered these rotation groups 1 through 100 (Table 2.3.6.3) according to the combination of the Patterson Laue groups. The characteristics of each of the 100 groups are given in Table 2.3.6.4, including the limits of the asymmetric unit. In the 100 unique combinations of noncubic Laue groups, there are only 16 basic rotation-function space groups.

2.3.6.4. Sampling, background and interpretation

If the origins are retained in the Pattersons, their product will form a high but constant plateau on which the rotation-function peaks are superimposed; this leads to a small apparent peak-to-noise ratio. The effect can be eliminated by removal of the origins through a modification of the Patterson coefficients. Irrespective of origin removal, a significant peak is one which is more than three r.m.s. deviations from the mean background.

As in all continuous functions sampled at discrete points, a convenient grid size must be chosen. Small intervals result in an excessive computing burden, while large intervals might miss peaks. Furthermore, equal increments of angles do not represent equal changes in rotation, which can result in distorted peaks (Lattman, 1972). In general, a crude idea of a useful sampling

Table 2.3.6.4. *Rotation-function Eulerian space groups*

The rotation space groups are given in Table 2.3.6.3.

No. of the rotation space group	No. of equivalent positions[(a)]	Symbol[(b)]	Translation along the θ_1 axis[(c)]	Translation along the θ_3 axis[(c)]	Range of the asymmetric unit[(d)]		
1	2	*Pn*	2π	2π	$0 \le \theta_1 < 2\pi$,	$0 \le \theta_2 \le \pi$,	$0 \le \theta_3 < 2\pi$
2	4	$Pbn2_1$	2π	2π	$0 \le \theta_1 < 2\pi$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < 2\pi$
3	4	*Pc*	π	2π	$0 \le \theta_1 < \pi$,	$0 \le \theta_2 \le \pi$,	$0 \le \theta_3 < 2\pi$
4	8	$Pbc2_1$	π	2π	$0 \le \theta_1 < \pi$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < 2\pi$
5	8	*Pc*	$\pi/2$	2π	$0 \le \theta_1 < \pi/2$,	$0 \le \theta_2 \le \pi$,	$0 \le \theta_3 < 2\pi$
6	16	$Pbc2_1$	$\pi/2$	2π	$0 \le \theta_1 < \pi/2$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < 2\pi$
7	6	*Pn*	$2\pi/3$	2π	$0 \le \theta_1 < 2\pi/3$,	$0 \le \theta_2 \le \pi$,	$0 \le \theta_3 < 2\pi$
8	12	$Pbn2_1$	$2\pi/3$	2π	$0 \le \theta_1 < 2\pi/3$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < 2\pi$
9	12	*Pc*	$\pi/3$	2π	$0 \le \theta_1 < \pi/3$,	$0 \le \theta_2 \le \pi$,	$0 \le \theta_3 < 2\pi$
10	24	$Pbc2_1$	$\pi/3$	2π	$0 \le \theta_1 < \pi/3$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < 2\pi$
11	4	$P2_1nb$	2π	2π	$0 \le \theta_1 < 2\pi$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < 2\pi$
12	8	*Pbnb*	2π	2π	$0 \le \theta_1 \le \pi/2$,	$0 \le \theta_2 < \pi$,	$0 \le \theta_3 < 2\pi$
13	8	*P2cb*	π	2π	$0 \le \theta_1 < \pi$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < 2\pi$
14	16	*Pbcb*	π	2π	$0 \le \theta_1 \le \pi/2$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < 2\pi$
15	16	*P2cb*	$\pi/2$	2π	$0 \le \theta_1 < \pi/2$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < 2\pi$
16	32	*Pbcb*	$\pi/2$	2π	$0 \le \theta_1 < \pi/2$,	$0 \le \theta_2 < \pi$,	$0 \le \theta_3 \le \pi/2$
17	12	$P2_1nb$	$2\pi/3$	2π	$0 \le \theta_1 < 2\pi/3$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < 2\pi$
18	24	*Pbnb*	$2\pi/3$	2π	$0 \le \theta_1 < 2\pi/3$,	$0 \le \theta_2 < \pi$,	$0 \le \theta_3 \le \pi/2$
19	24	*P2cb*	$\pi/3$	2π	$0 \le \theta_1 < \pi/3$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < 2\pi$
20	48	*Pbcb*	$\pi/3$	2π	$0 \le \theta_1 < \pi/3$,	$0 \le \theta_2 < \pi$,	$0 \le \theta_3 \le \pi/2$
21	4	*Pa*	2π	π	$0 \le \theta_1 < 2\pi$,	$0 \le \theta_2 \le \pi$,	$0 \le \theta_3 < \pi$
22	8	*Pba2*	2π	π	$0 \le \theta_1 < 2\pi$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < \pi$
23	8	*Pm*	π	π	$0 \le \theta_1 < \pi$,	$0 \le \theta_2 \le \pi$,	$0 \le \theta_3 < \pi$
24	16	*Pbm2*	π	π	$0 \le \theta_1 < \pi$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < \pi$
25	16	*Pm*	$\pi/2$	π	$0 \le \theta_1 < \pi/2$,	$0 \le \theta_2 \le \pi$,	$0 \le \theta_3 < \pi$
26	32	*Pbm2*	$\pi/2$	π	$0 \le \theta_1 < \pi/2$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < \pi$
27	12	*Pa*	$2\pi/3$	π	$0 \le \theta_1 < 2\pi/3$,	$0 \le \theta_2 \le \pi$,	$0 \le \theta_3 < \pi$
28	24	*Pba2*	$2\pi/3$	π	$0 \le \theta_1 < 2\pi/3$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < \pi$
29	24	*Pm*	$\pi/3$	π	$0 \le \theta_1 < \pi/3$,	$0 \le \theta_2 \le \pi$,	$0 \le \theta_3 < \pi$
30	48	*Pbm2*	$\pi/3$	π	$0 \le \theta_1 < \pi/3$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < \pi$
31	8	$P2_1ab$	2π	π	$0 \le \theta_1 < 2\pi$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < \pi$
32	16	*Pbab*	2π	π	$0 \le \theta_1 \le \pi/2$,	$0 \le \theta_2 < \pi$,	$0 \le \theta_3 < \pi$
33	16	*P2mb*	π	π	$0 \le \theta_1 < \pi$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < \pi$
34	32	*Pbmb*	π	π	$0 \le \theta_1 \le \pi/2$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < \pi$
35	32	*P2mb*	$\pi/2$	π	$0 \le \theta_1 < \pi/2$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < \pi$
36	64	*Pbmb*	$\pi/2$	π	$0 \le \theta_1 < \pi/2$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 \le \pi/2$
37	24	$P2_1ab$	$2\pi/3$	π	$0 \le \theta_1 < 2\pi/3$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < \pi$
38	48	*Pbab*	$2\pi/3$	π	$0 \le \theta_1 < 2\pi/3$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 \le \pi/2$
39	48	*P2mb*	$\pi/3$	π	$0 \le \theta_1 < \pi/3$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < \pi$
40	96	*Pbmb*	$\pi/3$	π	$0 \le \theta_1 < \pi/3$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 \le \pi/2$
41	8	*Pa*	2π	$\pi/2$	$0 \le \theta_1 < 2\pi$,	$0 \le \theta_2 \le \pi$,	$0 \le \theta_3 < \pi/2$
42	16	*Pba2*	2π	$\pi/2$	$0 \le \theta_1 < 2\pi$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < \pi/2$
43	16	*Pm*	π	$\pi/2$	$0 \le \theta_1 < \pi$,	$0 \le \theta_2 \le \pi$,	$0 \le \theta_3 < \pi/2$
44	32	*Pbm2*	π	$\pi/2$	$0 \le \theta_1 < \pi$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < \pi/2$
45	32	*Pm*	$\pi/2$	$\pi/2$	$0 \le \theta_1 < \pi/2$,	$0 \le \theta_2 \le \pi$,	$0 \le \theta_3 < \pi/2$
46	64	*Pbm2*	$\pi/2$	$\pi/2$	$0 \le \theta_1 < \pi/2$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < \pi/2$
47	24	*Pa*	$2\pi/3$	$\pi/2$	$0 \le \theta_1 < 2\pi/3$,	$0 \le \theta_2 \le \pi$,	$0 \le \theta_3 < \pi/2$
48	48	*Pba2*	$2\pi/3$	$\pi/2$	$0 \le \theta_1 < 2\pi/3$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < \pi/2$
49	48	*Pm*	$\pi/3$	$\pi/2$	$0 \le \theta_1 < \pi/3$,	$0 \le \theta_2 \le \pi$,	$0 \le \theta_3 < \pi/2$
50	96	*Pbm2*	$\pi/3$	$\pi/2$	$0 \le \theta_1 < \pi/3$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < \pi/2$
51	16	$P2_1ab$	2π	$\pi/2$	$0 \le \theta_1 < 2\pi$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < \pi/2$
52	32	*Pbab*	2π	$\pi/2$	$0 \le \theta_1 < 2\pi$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 \le \pi/4$
53	32	*P2mb*	π	$\pi/2$	$0 \le \theta_1 < \pi$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < \pi/2$
54	64	*Pbmb*	π	$\pi/2$	$0 \le \theta_1 \le \pi/2$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < \pi/2$
55	64	*P2mb*	$\pi/2$	$\pi/2$	$0 \le \theta_1 < \pi/2$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < \pi/2$
56	128	*Pbmb*	$\pi/2$	$\pi/2$	$0 \le \theta_1 \le \pi/4$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < \pi/2$
57	48	$P2_1ab$	$2\pi/3$	$\pi/2$	$0 \le \theta_1 < 2\pi/3$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < \pi/2$
58	96	*Pbab*	$2\pi/3$	$\pi/2$	$0 \le \theta_1 < 2\pi/3$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 \le \pi/4$

Table 2.3.6.4 (*cont.*)

No. of the rotation space group	No. of equivalent positions[(a)]	Symbol[(b)]	Translation along the θ_1 axis[(c)]	Translation along the θ_3 axis[(c)]	Range of the asymmetric unit[(d)]		
59	96	*P2mb*	$\pi/3$	$\pi/2$	$0 \le \theta_1 < \pi/3$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < \pi/2$
60	192	*Pbmb*	$\pi/3$	$\pi/2$	$0 \le \theta_1 \le \pi/6$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < \pi/2$
61	6	*Pn*	2π	$2\pi/3$	$0 \le \theta_1 < 2\pi$,	$0 \le \theta_2 \le \pi$,	$0 \le \theta_3 < 2\pi/3$
62	12	$Pbn2_1$	2π	$2\pi/3$	$0 \le \theta_1 < 2\pi$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < 2\pi/3$
63	12	*Pc*	π	$2\pi/3$	$0 \le \theta_1 < \pi$,	$0 \le \theta_2 \le \pi$,	$0 \le \theta_3 < 2\pi/3$
64	24	$Pbc2_1$	π	$2\pi/3$	$0 \le \theta_1 < \pi$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < 2\pi/3$
65	24	*Pc*	$\pi/2$	$2\pi/3$	$0 \le \theta_1 < \pi/2$,	$0 \le \theta_2 \le \pi$,	$0 \le \theta_3 < 2\pi/3$
66	48	$Pbc2_1$	$\pi/2$	$2\pi/3$	$0 \le \theta_1 < \pi/2$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < 2\pi/3$
67	18	*Pn*	$2\pi/3$	$2\pi/3$	$0 \le \theta_1 < 2\pi/3$,	$0 \le \theta_2 \le \pi$,	$0 \le \theta_3 < 2\pi/3$
68	36	$Pbn2_1$	$2\pi/3$	$2\pi/3$	$0 \le \theta_1 < 2\pi/3$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < 2\pi/3$
69	36	*Pc*	$\pi/3$	$2\pi/3$	$0 \le \theta_1 < \pi/3$,	$0 \le \theta_2 \le \pi$,	$0 \le \theta_3 < 2\pi/3$
70	72	$Pbc2_1$	$\pi/3$	$2\pi/3$	$0 \le \theta_1 < \pi/3$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < 2\pi/3$
71	12	$P2_1nb$	2π	$2\pi/3$	$0 \le \theta_1 < 2\pi$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < 2\pi/3$
72	24	*Pbnb*	2π	$2\pi/3$	$0 \le \theta_1 \le \pi/2$,	$0 \le \theta_2 < \pi$,	$0 \le \theta_3 < 2\pi/3$
73	24	*P2cb*	π	$2\pi/3$	$0 \le \theta_1 < \pi$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < 2\pi/3$
74	48	*Pbcb*	π	$2\pi/3$	$0 \le \theta_1 \le \pi/2$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < 2\pi/3$
75	48	*P2cb*	$\pi/2$	$2\pi/3$	$0 \le \theta_1 < \pi/2$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < 2\pi/3$
76	96	*Pbcb*	$\pi/2$	$2\pi/3$	$0 \le \theta_1 \le \pi/4$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < 2\pi/3$
77	36	$P2_1nb$	$2\pi/3$	$2\pi/3$	$0 \le \theta_1 < 2\pi/3$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < 2\pi/3$
78	72	*Pbnb*	$2\pi/3$	$2\pi/3$	$0 \le \theta_1 \le \pi/6$,	$0 \le \theta_2 \le \pi$,	$0 \le \theta_3 < 2\pi/3$
79	72	*P2 cb*	$\pi/3$	$2\pi/3$	$0 \le \theta_1 < \pi/3$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < 2\pi/3$
80	144	*Pbcb*	$\pi/3$	$2\pi/3$	$0 \le \theta_1 \le \pi/6$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < 2\pi/3$
81	12	*Pa*	2π	$\pi/3$	$0 \le \theta_1 < 2\pi$,	$0 \le \theta_2 \le \pi$,	$0 \le \theta_3 < \pi/3$
82	24	*Pba2*	2π	$\pi/3$	$0 \le \theta_1 < 2\pi$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < \pi/3$
83	24	*Pm*	π	$\pi/3$	$0 \le \theta_1 < \pi$,	$0 \le \theta_2 \le \pi$,	$0 \le \theta_3 < \pi/3$
84	48	*Pbm2*	π	$\pi/3$	$0 \le \theta_1 < \pi$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < \pi/3$
85	48	*Pm*	$\pi/2$	$\pi/3$	$0 \le \theta_1 < \pi/2$,	$0 \le \theta_2 \le \pi$,	$0 \le \theta_3 < \pi/3$
86	96	*Pbm2*	$\pi/2$	$\pi/3$	$0 \le \theta_1 < \pi/2$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < \pi/3$
87	36	*Pa*	$2\pi/3$	$\pi/3$	$0 \le \theta_1 < 2\pi/3$,	$0 \le \theta_2 \le \pi$,	$0 \le \theta_3 < \pi/3$
88	72	*Pba2*	$2\pi/3$	$\pi/3$	$0 \le \theta_1 < 2\pi/3$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < \pi/3$
89	72	*Pm*	$\pi/3$	$\pi/3$	$0 \le \theta_1 < \pi/3$,	$0 \le \theta_2 \le \pi$,	$0 \le \theta_3 < \pi/3$
90	144	*Pbm2*	$\pi/3$	$\pi/3$	$0 \le \theta_1 < \pi/3$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < \pi/3$
91	24	$P2_1ab$	2π	$\pi/3$	$0 \le \theta_1 < 2\pi$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < \pi/3$
92	48	*Pbab*	2π	$\pi/3$	$0 \le \theta_1 \le \pi/2$,	$0 \le \theta_2 < \pi$,	$0 \le \theta_3 < \pi/3$
93	48	*P2mb*	π	$\pi/3$	$0 \le \theta_1 < \pi$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < \pi/3$
94	96	*Pbmb*	π	$\pi/3$	$0 \le \theta_1 \le \pi/2$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 \le \pi/2$
95	96	*P2mb*	$\pi/2$	$\pi/3$	$0 \le \theta_1 < \pi/2$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < \pi/3$
96	192	*Pbmb*	$\pi/2$	$\pi/3$	$0 \le \theta_1 \le \pi/4$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < \pi/3$
97	72	$P2_1ab$	$2\pi/3$	$\pi/3$	$0 \le \theta_1 < 2\pi/3$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < \pi/3$
98	144	*Pbab*	$2\pi/3$	$\pi/3$	$0 \le \theta_1 < 2\pi/3$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 \le \pi/6$
99	144	*P2mb*	$\pi/3$	$\pi/3$	$0 \le \theta_1 < \pi/3$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < \pi/3$
100	288	*Pbmb*	$\pi/3$	$\pi/3$	$0 \le \theta_1 \le \pi/6$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < \pi/3$

Notes: (*a*) This is the number of equivalent positions in the rotation unit cell. (*b*) Each symbol retains the order $\theta_1, \theta_2, \theta_3$. The monoclinic space groups have the *b* axis unique setting. (*c*) This is a translation symmetry: *e.g.* for the case of $\pi/2$ translation along the θ_1 axis, $\theta_1, \theta_2, \theta_3$ goes to $\pi/2 + \theta_1, \theta_2, \theta_3$ and $\pi + \theta_1, \theta_2, \theta_3$, and $3\pi/2 + \theta_1, \theta_2, \theta_3$. All other equivalent positions in the basic rotation space group are similarly translated. (*d*) Several consistent sets of ranges exist but the one with the minimum range of θ_2 is listed.

interval can be obtained by considering the angle necessary to move one reciprocal-lattice point onto its neighbour (separated by a^*) at the extremity of the resolution limit, R. This interval is given by

$$\Delta\theta = a^*/2(1/R) = \tfrac{1}{2}Ra^*.$$

Simple sharpening of the rotation function can be useful. This can be achieved by restricting the computations to a shell in reciprocal space or by using normalized structure factors. Useful limits are frequently 10 to 6, 10 to 4 or 10 to 3.5 Å for an average protein or 6 to 5 Å for a virus structure determination. In addition, use of restricted resolution ranges, such as 6 to 5 Å or 3.5 to 3.0 Å, has been found in numerous cases to give especially well defined results (Arnold *et al.*, 1984).

When exploring the rotation function in polar coordinates, there is no significance to the latitude φ (Fig. 2.3.6.4) when $\psi = 0$. For small values of ψ, the rotation function will be quite insensitive to φ, which therefore needs to be explored only at coarse intervals (say 45°). As ψ approaches the equator at 90°, optimal increments of ψ and φ will be about equal. A similar situation exists with Eulerian angles. When $\theta_2 = 0$, the rotation function will be determined by $\theta_1 + \theta_3$, corresponding to $\psi = 0$ and

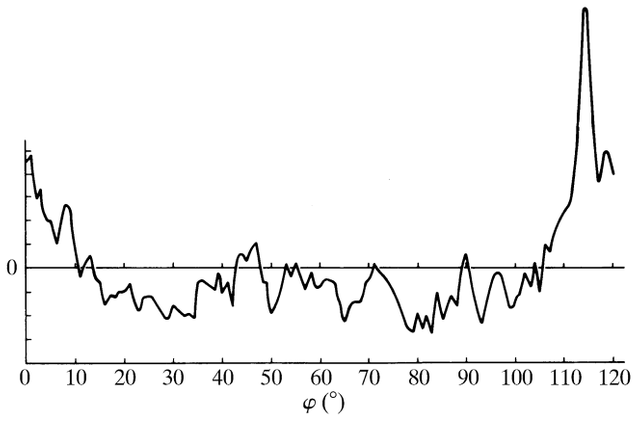


Fig. 2.3.6.6. The locked rotation function, L, applied to the determination of the orientation of the common cold virus (Arnold *et al.*, 1984). There are four virus particles per cubic cell with each particle sitting on a threefold axis. The locked rotation function explores all positions of rotation about this axis and, hence, repeats itself after 120°. The locked rotation function is determined from the individual rotation-function values of the noncrystallographic symmetry directions of a 532 icosahedron. [Reprinted with permission from Arnold *et al.* (1984).]

varying κ in polar coordinates. There will be no dependence on $(\theta_1 - \theta_3)$. Thus Eulerian searches can often be performed more economically in terms of the variables $\eta = \theta_1 + \theta_3$ and $\Delta = \theta_1 - \theta_3$, where

$$[\boldsymbol{\rho}] = \begin{pmatrix} \left[\cos\eta\cos^2\left(\frac{\theta_2}{2}\right) + \cos\Delta\sin^2\left(\frac{\theta_2}{2}\right)\right] & \left[\sin\eta\cos^2\left(\frac{\theta_2}{2}\right) + \sin\Delta\sin^2\left(\frac{\theta_2}{2}\right)\right] & \sin\theta_2\sin(\eta-\Delta) \\ \left[-\sin\eta\cos^2\left(\frac{\theta_2}{2}\right) + \sin\Delta\sin^2\left(\frac{\theta_2}{2}\right)\right] & \left[\cos\eta\cos^2\left(\frac{\theta_2}{2}\right) - \cos\Delta\sin^2\left(\frac{\theta_2}{2}\right)\right] & \sin\theta_2\cos(\eta-\Delta) \\ \sin\theta_2\sin(\eta+\Delta) & -\sin\theta_2\cos(\eta+\Delta) & \cos\theta_2 \end{pmatrix},$$

which reduces to the simple rotation matrix

$$[\boldsymbol{\rho}] = \begin{pmatrix} \cos\eta & \sin\eta & 0 \\ -\sin\eta & \cos\eta & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

when $\theta_2 = 0$.

The computational effort to explore carefully a complete asymmetric unit of the rotation-function Eulerian group can be considerable. However, unless improper rotations are under investigation (as, for example, cross-rotation functions between different crystal forms of the same molecule), it is not generally necessary to perform such a global search. The number of molecules per crystallographic asymmetric unit, or the number of subunits per molecule, are often good indicators as to the possible types of noncrystallographic symmetry element. For instance, in the early investigation of insulin, the rotation function was used to explore only the $\kappa = 180°$ plane in polar coordinates as there were only two molecules per crystallographic asymmetric unit (Dodson *et al.*, 1966). Rotation functions of viruses, containing 532 icosahedral symmetry, are usually limited to exploration of the κ = 180, 120, 72 and 144° planes [*e.g.* Rayment *et al.* (1978) and Arnold *et al.* (1984)].

In general, the interpretation of the rotation function is straightforward. However, noise often builds up relative to the signal in high-symmetry space groups or if the data are limited or poor. One aid to a systematic interpretation is the locked rotation function (Rossmann *et al.*, 1972) for use when a molecule has more than one noncrystallographic symmetry axis. It is then possible to determine the rotation-function values for each molecular axis for a chosen molecular orientation (Fig. 2.3.6.6) (see Section 2.3.6.6).

Another problem in the interpretation of rotation functions is the appearance of apparent noncrystallographic symmetry that relates the self-Patterson of one molecule to the self-Patterson of a crystallographically related molecule. For example, take the case of α-chymotrypsin (Blow *et al.*, 1964). The space group is $P2_1$ with a molecular dimer in each of the two crystallographic asymmetric units. The noncrystallographic dimer axis was found to be perpendicular to the crystallographic 2_1 axis. The product of the crystallographic twofold in the Patterson with the orthogonal twofold in the self-Patterson vectors around the origin creates a third twofold, orthogonal to both of the other twofolds. In real space this represents a twofold screw direction relating the two dimers in the cell. In other cases, the product of the crystallographic and noncrystallographic symmetry results in symmetry which only has meaning in terms of all the vectors in the vicinity of the Patterson origin, but not in real space. Rotation-function peaks arising from such products are called Klug peaks (Johnson *et al.*, 1975). Such peaks normally refer to the total symmetry of all the vectors around the Patterson origin and may, therefore, be much larger than the peaks due to noncrystallographic symmetry within one molecule alone. Hence the Klug peaks, if not correctly recognized, can lead to erroneous conclusions (Åkervall *et al.*, 1972). Litvin (1975) has shown how Klug peaks can be predicted. These usually occur only for special orientations of a particle with a given symmetry relative to the crystallographic symmetry axes. Prediction of Klug peaks requires the simultaneous consideration of the noncrystallographic point group, the crystallographic point group and their relative orientations.

A special, but frequently occurring, situation arises when an evenfold noncrystallographic symmetry operator (*e.g.* 2-, 4-, 6-, 8-*etc.* fold axes) is parallel, or nearly parallel, to a crystallographic evenfold axis or screw axis. If the crystallographic evenfold axis is, say, parallel to Z, then if the centre of molecule I is at (X_0, Y_0, Z_0), the centre of molecule II will be at $(-X_0, -Y_0, Z_0)$. If molecule I has an evenfold axis parallel to Z, then for every atom (a) at $(x + X_0, y + Y_0, z + Z_0)$, there will be an atom (b) at $(-x + X_0, -y + Y_0, z + Z_0)$. The crystallographic symmetry-equivalent positions of these two atoms in molecule II will be at (c) $(-x - X_0, -y - Y_0, z + Z_0)$ and (d) $(x - X_0, y - Y_0, z + Z_0)$. The vectors between atoms (a) and (d) and also between atoms (b) and (c) will both have component lengths of $(2X_0, 2Y_0, 0)$. The position of this vector in a Patterson map is independent of the actual atoms in the molecule and depends only on the position of the molecular noncrystallographic symmetry axis.

Every atom will produce two vectors of this type, all of which will accumulate in a Patterson map to produce a large peak, which establishes the exact position of the noncrystallographic symmetry evenfold axis relative to the crystallographic axis. The position of the special peak is on the Harker section, namely at w = 0 for a crystallographic twofold axis and at $w = \frac{1}{2}$ for a crystallographic 2_1 screw axis. If there are N atoms in the structure of the two crystallographically related dimers, then the height of the origin is proportional to N (the number of zero-length vectors). The number of vectors with length $(2X_0, 2Y_0, 0)$ will be twice the number of atoms in each monomer, or $2 \times (N/4)$, which is $N/2$. Thus the special peak should be about half the height of the Patterson map's origin peak. In practice, the peak is often somewhat lower because the noncrystallographic symmetry and crystallographic axes might not be exactly parallel. This situation can be mitigated by computing the Patterson map with lower-resolution reflections only, as the difference in orientation between the axes is less significant when viewed at lower resolution (McKenna *et al.*, 1992).

2.3.6.5. The fast rotation function

Unfortunately, the rotation-function computations can be extremely time-consuming by conventional methods. Sasada (1964) developed a technique for rapidly finding the maximum of a given peak by looking at the slope of the rotation function. A major breakthrough came when Crowther (1972) recast the rotation function in a manner suitable for rapid computation. Only a brief outline of Crowther's fast rotation function is given here. Details are found in the original text (Crowther, 1972) and his computer program description.

Since the rotation function correlates spherical volumes of a given Patterson density with rotated versions of either itself or another Patterson density, it is likely that a more natural form for the rotation function will involve spherical harmonics rather than the Fourier components $|\mathbf{F_h}|^2$ of the crystal representation. Thus, if the two Patterson densities $P_1(r, \psi, \varphi)$ and $P_2(r, \psi, \varphi)$ are expanded within the spherical volume of radius less than a limiting value of a, then

$$P_1(r, \psi, \varphi) = \sum_{lmn} a^*_{lmn} \hat{j}_l(k_{ln} r) \hat{Y}_l^{m^*}(\psi, \varphi)$$

and

$$P_2(r, \psi, \varphi) = \sum_{l'm'n'} b_{l'm'n'} \hat{j}_{l'}(k_{l'n'} r) \hat{Y}_{l'}^{m'}(\psi, \varphi),$$

and the rotation function would then be defined as

$$R = \int_{\text{sphere}} P_1(r, \psi, \varphi) \mathscr{R} P_2(r, \psi, \varphi) r^2 \sin\psi \, \mathrm{d}r \, \mathrm{d}\psi \, \mathrm{d}\varphi.$$

Here $\hat{Y}_l^m(\psi, \varphi)$ is the normalized spherical harmonic of order l; $\hat{j}_l(k_{ln} r)$ is the normalized spherical Bessel function of order l; a_{lmn}, b_{lmn} are complex coefficients; and $\mathscr{R}P_2(r, \psi, \varphi)$ represents the rotated second Patterson. The rotated spherical harmonic can then be expressed in terms of the Eulerian angles $\theta_1, \theta_2, \theta_3$ as

$$\mathscr{R}(\theta_1, \theta_2, \theta_3) \hat{Y}_l^m(\psi, \varphi) = \sum_{q=-l}^{l} D^l_{qm}(\theta_1, \theta_2, \theta_3) \hat{Y}_l^q(\psi, \varphi),$$

where

$$D^l_{qm}(\theta_1, \theta_2, \theta_3) = \exp(iq\theta_3) d^l_{qm}(\theta_2) \exp(im\theta_1)$$

and $d^l_{qm}(\theta_2)$ are the matrix elements of the three-dimensional rotation group. It can then be shown that

$$R(\theta_1, \theta_2, \theta_3) = \sum_{lmm'n} a^*_{lmn} b_{lm'n} D^l_{m'm}(\theta_1, \theta_2, \theta_3).$$

Since the radial summation over n is independent of the rotation,

$$c_{lmm'} = \sum_n a^*_{lmn} b_{lmn},$$

and hence

$$R(\theta_1, \theta_2, \theta_3) = \sum_{lmm'} c_{lmm'} D^l_{m'm}(\theta_1, \theta_2, \theta_3)$$

or

$$R(\theta_1, \theta_2, \theta_3) = \sum_{mm'} \left[\sum_l c_{lmm'} d^l_{m'm}(\theta_2) \right] \exp[i(m'\theta_3 + m\theta_1)].$$

The coefficients $c_{lmm'}$ refer to a particular pair of Patterson densities and are independent of the rotation. The coefficients $D^l_{m'm}$, containing the whole rotational part, refer to rotations of spherical harmonics and are independent of the particular Patterson densities. Since the summations over m and m' represent a Fourier synthesis, rapid calculation is possible.

As polar coordinates rather than Eulerian angles provide a more graphic interpretation of the rotation function, Tanaka (1977) has recast the initial definition as

$$\begin{aligned} R(\theta_1, \theta_2, \theta_3) = & \int_{\text{sphere}} [\mathscr{R}(\theta_1, \theta_2, \theta_3 = 0) P_1(r, \psi, \varphi)] \\ & \times [\mathscr{R}(\theta_1, \theta_2, \theta_3) P_2(r, \psi, \varphi)] \, \mathrm{d}V \\ = & \int_{\text{sphere}} [P_1(r, \psi, \varphi)][\mathscr{R}^{-1}(\theta_1, \theta_2, \theta_3 = 0) \\ & \times \mathscr{R}(\theta_1, \theta_2, \theta_3) P_2(r, \psi, \varphi)] \, \mathrm{d}V. \end{aligned}$$

He showed that the polar coordinates are now equivalent to $\kappa = \theta_3$, $\psi = \theta_2$ and $\varphi = \theta_1 - \pi/2$. The rotation function can then be expressed as

$$\begin{aligned} R(\kappa, \psi, \varphi) = \sum_{lmm'} \left(\sum_n a^*_{lmn} b_{lm'n} \right) \sum_q \{ & d^l_{qm}(\psi) d^l_{qm'}(\psi) (-1)^{(m'-m)} \\ & \times \exp[i(\kappa q)] \exp[i(m' - m)\varphi] \}, \end{aligned}$$

permitting rapid calculation of the fast rotation function in polar coordinates.

Crowther (1972) uses the Eulerian angles α, β, γ which are related to those defined by Rossmann & Blow (1962) according to $\theta_1 = \alpha + \pi/2$, $\theta_2 = \beta$ and $\theta_3 = \gamma - \pi/2$.

An alternative formulation of the fast rotation function, which reduces the errors in the calculation, is implemented in *AMoRe* (Navaza, 1987, 1993, 1994, 2001*a*). New target functions derived from the principle of maximum likelihood have been implemented in conjunction with fast rotation functions in the program *Phaser*, which can also take advantage of partial model information in orienting unknown fragments (Storoni *et al.*, 2004).

2.3.6.6. Locked rotation functions

Many oligomers of macromolecules obey simple point-group symmetry, which is maintained as noncrystallographic symmetry when they are crystallized. For example, a homo-tetramer often obeys 222 point-group symmetry, and icosahedral viruses obey 532 point-group symmetry. The locked rotation function takes advantage of this information and can greatly simplify the calculation and the interpretation of rotation functions (Fig. 2.3.6.6) (Arnold *et al.*, 1984; Rossmann *et al.*, 1972; Tong, 2001*a*; Tong & Rossmann, 1990, 1997). During the rotation-function calculation, the noncrystallographic symmetry of the crystal is *locked* to the presumed point group, hence the name locked rotation function.

Given the noncrystallographic symmetry point group, a standard orientation can be defined which serves as a reference orientation for this point group. For example, for 222 point-group symmetry, the standard orientation can be defined such that the three twofold axes are parallel to the three Cartesian coordinate axes that are defined with respect to the crystal unit cell. Once the standard orientation is defined, any orientation of the noncrystallographic symmetry point group can be related to the standard

orientation by a single set of three rotation angles that determine the rotation matrix [**E**].

Assume $[\mathbf{I}_n]$ ($n = 1, \ldots, N$) is the collection of noncrystallographic symmetry point-group rotation matrices in the standard orientation. Then the operation [**E**] will bring the noncrystallographic symmetry point group to a new orientation and the noncrystallographic symmetry rotation matrices in this new orientation, $[\boldsymbol{\rho}_n]$, are given by (Tong & Rossmann, 1990)

$$[\boldsymbol{\rho}_n] = [\mathbf{E}][\mathbf{I}_n][\mathbf{E}]^{-1}. \qquad (2.3.6.8)$$

For each rotation [**E**], the ordinary self-rotation-function value (R_n) for each of the noncrystallographic symmetry rotation matrices in the new orientation ($[\boldsymbol{\rho}_n]$) is calculated. The locked self-rotation-function value ($R_{\rm L}$) for this rotation is defined as the average of the ordinary rotation-function values over the noncrystallographic symmetry elements

$$R_{\rm L}([\mathbf{E}]) = \frac{1}{N-1}\sum_{n=2}^{N} R_n,$$

where the summation starts from 2 as it is assumed that $[\mathbf{I}_1]$ is the identity matrix.

The locked self rotation function simplifies the task of interpreting the self rotation function for the orientation of an noncrystallographic symmetry assembly. Instead of searching for $N-1$ peaks in the ordinary self rotation function, a single peak is sought in the locked self rotation function. It must be emphasized that this rotation ([**E**]) in the locked self rotation function is most often a general rotation. The locked self rotation function also reduces the noise in the rotation-function calculation by a factor of $(N-1)^{1/2}$ due to the averaging of the ordinary rotation-function values (Tong & Rossmann, 1990).

The symmetry of the locked self rotation function is generally rather complex and an analytical solution is often impossible (Tong & Rossmann, 1990). It depends not only on the crystallographic symmetry and the noncrystallographic symmetry, but also on the definition of the standard orientation of the noncrystallographic symmetry. For example, if the standard orientation is defined such that the twofold axes are parallel to the Cartesian coordinate axes for the 222 point group, a 90° rotation around the X, Y or Z axis, or a 120° rotation around the 111 direction, does not cause a net change to the standard orientation. Such rotations will appear as symmetry in the locked self rotation function (Tong & Rossmann, 1997). In practice, the locked self rotation function can be calculated rather quickly, especially if the fast rotation function is used. A large region of rotation space can be explored in the calculation of the locked rotation function and the solutions can then be clustered based on the resulting orientation of the noncrystallographic symmetry. For example, two rotations $[\mathbf{E}_1]$ and $[\mathbf{E}_2]$ that produce the same set of noncrystallographic symmetry matrices based on (2.3.6.8) are likely to be related by the symmetry of the locked self rotation function.

A locked cross rotation function can also be defined to determine the orientation, [**F**], of the known monomer structure relative to the noncrystallographic symmetry of the molecular assembly (Navaza *et al.*, 1998; Tong, 2001*a*; Tong & Rossmann, 1990, 1997). With the knowledge of [**F**] and the orientation of the noncrystallographic symmetry in the crystal [**E**], which can be determined from the locked self rotation function, the orientation of all the monomers in the crystal cell is given by

$$[\boldsymbol{\rho}_n] = [\mathbf{E}][\mathbf{I}_n][\mathbf{F}].$$

Therefore, $[\boldsymbol{\rho}_n]$ represents the rotational relationship between the monomer search model and the monomers of the assembly in the crystal. An ordinary cross-rotation-function value R_n can be calculated for each of the rotations $[\boldsymbol{\rho}_n]$, and the locked cross-rotation-function value is defined as the average

$$R_{\rm L} = (1/N)\sum_n R_n.$$

Like the locked self rotation function, the locked cross rotation function can determine the orientation of all the monomers of the noncrystallographic symmetry assembly with a single rotation.

2.3.7. Translation functions

2.3.7.1. Introduction

The problem of determining the position of a noncrystallographic symmetry element in space, or the position of a molecule of known orientation in a unit cell, has been reviewed by Rossmann (1972), Colman *et al.* (1976), Karle (1976), Argos & Rossmann (1980), Harada *et al.* (1981) and Beurskens (1981). All methods depend on the prior knowledge of the object's orientation implied by the rotation matrix [***C***]. The various translation functions, T, derived below, can only be computed given this information.

The general translation function can be defined as

$$T(\mathbf{S}_x, \mathbf{S}_{x'}) = \int_U \rho_1(\mathbf{x}) \cdot \rho_2(\mathbf{x}')\, {\rm d}\mathbf{x},$$

where T is a six-variable function given by each of the three components that define $\mathbf{S}_x$ and $\mathbf{S}_{x'}$. Here $\mathbf{S}_x$ and $\mathbf{S}_{x'}$ are equivalent reference positions of the objects, whose densities are $\rho_1(\mathbf{x})$ and $\rho_2(\mathbf{x}')$. The translation function searches for the optimal overlap of the two objects after they have been similarly oriented. Following the same procedure used for the rotation-function derivation, Fourier summations are substituted for $\rho_1(\mathbf{x})$ and $\rho_2(\mathbf{x}')$. It can then be shown that

$$T(\mathbf{S}_x, \mathbf{S}_{x'}) = \int_U \left\{\frac{1}{V_{\mathbf{h}}}\sum_{\mathbf{h}} |\mathbf{F}_{\mathbf{h}}| \exp[i(\alpha_{\mathbf{h}} - 2\pi\mathbf{h}\cdot\mathbf{x})]\right\} \times \left\{\frac{1}{V_{\mathbf{p}}}\sum_{\mathbf{p}} |\mathbf{F}_{\mathbf{p}}| \exp[i(\alpha_{\mathbf{p}} - 2\pi\mathbf{p}\cdot\mathbf{x}')]\right\} {\rm d}\mathbf{x}.$$

Using the substitution $\mathbf{x}' = [\boldsymbol{C}]\mathbf{x} + \mathbf{d}$ and simplifying leads to

$$T(\mathbf{S}_x, \mathbf{S}'_x) = \frac{1}{V_{\mathbf{h}}V_{\mathbf{p}}}\sum_{\mathbf{h}}\sum_{\mathbf{p}} |\mathbf{F}_{\mathbf{h}}||\mathbf{F}_{\mathbf{p}}| \times \exp[i(\alpha_{\mathbf{h}} + \alpha_{\mathbf{p}} - 2\pi\mathbf{p}\cdot\mathbf{d})] \times \int_U \exp\{-2\pi i(\mathbf{h} + [\boldsymbol{C}]^T\mathbf{p})\cdot\mathbf{x}\}\, {\rm d}\mathbf{x}.$$

The integral is the diffraction function $G_{\mathbf{hp}}$ (2.3.6.4). If the integration is taken over the volume U, centred at $\mathbf{S}_x$ and $\mathbf{S}_{x'}$, it follows that

$$T(\mathbf{S}_x, \mathbf{S}_{x'}) = \frac{2}{V_{\mathbf{h}}V_{\mathbf{p}}}\sum_{\mathbf{h}}\sum_{\mathbf{p}} |\mathbf{F}_{\mathbf{h}}||\mathbf{F}_{\mathbf{p}}| G_{\mathbf{hp}} \times \cos[\alpha_{\mathbf{h}} + \alpha_{\mathbf{p}} - 2\pi(\mathbf{h}\cdot\mathbf{S}_x + \mathbf{p}\cdot\mathbf{S}_{x'})]. \qquad (2.3.7.1)$$

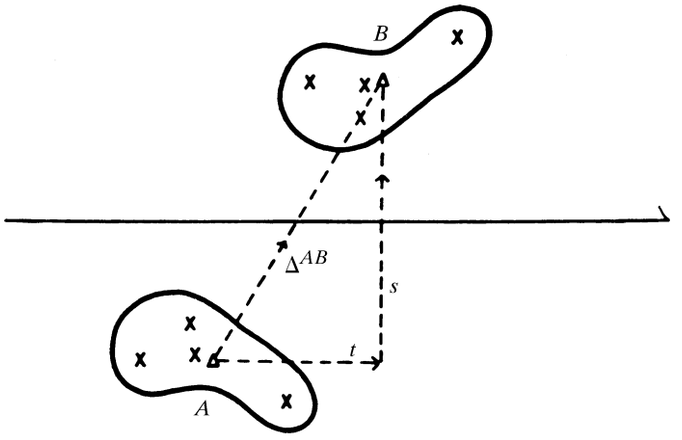


Fig. 2.3.7.1. Crosses represent atoms in a two-dimensional model structure. The triangles are the points chosen as approximate centres of molecules A and B. Δ^{AB} has components t and s parallel and perpendicular, respectively, to the screw rotation axis. [Reprinted from Rossmann *et al.* (1964).]

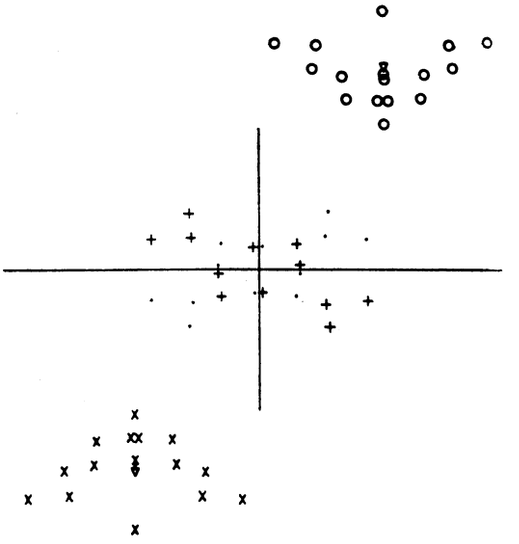


Fig. 2.3.7.2. Vectors arising from the structure in Fig. 2.3.7.1. The self-vectors of molecules A and B are represented by + and ·; the cross-vectors from molecules A to B and B to A by × and ○. Triangles mark the position of $+\Delta^{AB}$ and $-\Delta^{AB}$. [Reprinted from Rossmann *et al.* (1964).]

2.3.7.2. Position of a noncrystallographic element relating two unknown structures

The function (2.3.7.1) is quite general. For instance, the rotation function corresponds to a comparison of Patterson functions P_1 and P_2 at their origins. That is, the coefficients are F^2, phases are zero and $\mathbf{S}_x = \mathbf{S}_{x'} = 0$. However, the determination of the translation between two objects requires the comparison of cross-vectors away from the origin.

Consider, for instance, the determination of the precise translation vector parallel to a rotation axis between two identical molecules of unknown structure. For simplicity, let the noncrystallographic axis be a dyad (Fig. 2.3.7.1). Fig. 2.3.7.2 shows the corresponding Patterson of the hypothetical point-atom structure. Opposite sets of cross-Patterson vectors in Fig. 2.3.7.2 are related by a twofold rotation and a translation equal to twice the precise vector in the original structure. A suitable translation function would then compare a Patterson at $\mathbf{S}$ with the rotated Patterson at $-\mathbf{S}$. Hence, substituting $\mathbf{S}_x = \mathbf{S}$ and $\mathbf{S}_{x'} = -\mathbf{S}$ in (2.3.7.1),

$$T(\mathbf{S}) = \frac{2}{V^2}\sum_{\mathbf{h}}\sum_{\mathbf{p}}|\mathbf{F}_\mathbf{h}|^2|\mathbf{F}_\mathbf{p}|^2 G_{\mathbf{hp}}\cos[2\pi(\mathbf{h}-\mathbf{p})\cdot\mathbf{S}]. \qquad (2.3.7.2)$$

The opposite cross-vectors can be superimposed only if an evenfold rotation between the unknown molecules exists. The translation function (2.3.7.2) is thus applicable only in this special situation. There is no published translation method to determine the interrelation of two unknown structures in a crystallographic asymmetric unit or in two different crystal forms. However, another special situation exists if a molecular evenfold axis is parallel to a crystallographic evenfold axis. In this case, the position of the noncrystallographic symmetry element can be easily determined from the large peak in the corresponding Harker section of the Patterson.

In general, it is difficult or impossible to determine the positions of noncrystallographic axes (or their intersection at a molecular centre). However, the position of heavy atoms in isomorphous derivatives, which usually obey the noncrystallographic symmetry, can often determine this information.

2.3.7.3. Position of a known molecular structure in an unknown unit cell

The most common type of translation function occurs when looking for the position of a known molecular structure in an unknown crystal. For instance, if the structure of an enzyme has previously been determined by the isomorphous replacement method, then the structure of the same enzyme from another species can often be solved by molecular replacement [*e.g.* Grau *et al.* (1981)]. However, there are some severe pitfalls when, for instance, there are gross conformational changes [*e.g.* Moras *et al.* (1980)]. This type of translation function could also be useful in the interpolation of E maps produced by direct methods. Here there may often be confusion as a consequence of a number of molecular images related by translations (Karle, 1976; Beurskens, 1981; Egert & Sheldrick, 1985).

Tollin's (1966) Q function and Crowther & Blow's (1967) translation function are essentially identical (Tollin, 1969) and depend on a prior knowledge of the search molecule as well as its orientation in the unknown cell. The derivation given here, however, is somewhat more general and follows the derivation of Argos & Rossmann (1980), and should be compared with the method of Harada *et al.* (1981).

If the known molecular structure is correctly oriented into a cell (**p**) of an unknown structure and placed at **S** with respect to a defined origin, then a suitable translation function is

$$T(\mathbf{S}) = \sum_{\mathbf{p}}|\mathbf{F}_{\mathbf{p},\,\mathrm{obs}}|^2|\mathbf{F}_\mathbf{p}(\mathbf{S})|^2. \qquad (2.3.7.3)$$

This definition is preferable to one based on an R-factor calculation as it is more amenable to computation and is independent of a relative scale factor.

The structure factor $\mathbf{F}_\mathbf{p}(\mathbf{S})$ can be calculated by modifying expression (2.3.8.9) (see below). That is,

$$\mathbf{F}_\mathbf{p}(\mathbf{S}) = \frac{U}{V_\mathbf{h}}\sum_{n=1}^{N}\exp(2\pi i\mathbf{p}\cdot\mathbf{S}_n)\left[\sum_{\mathbf{h}}\mathbf{F}_\mathbf{h}G_{\mathbf{hp}_n}\exp(-2\pi i\mathbf{h}\cdot\mathbf{S})\right],$$

where $V_\mathbf{h}$ is the volume of cell (**h**) and $\mathbf{S}_n$ is the position, in the nth crystallographic asymmetric unit, of cell (**p**) corresponding to **S** in known cell (**h**). Let

$$A_{\mathbf{p},\,n}\exp(i\gamma_n) = \sum_{\mathbf{h}}\mathbf{F}_\mathbf{h}G_{\mathbf{hp}_n}\exp(-2\pi i\mathbf{h}\cdot\mathbf{S}),$$

which are the coefficients of the molecular transform for the known molecule placed into the nth asymmetric unit of the **p** cell. Thus

$$\mathbf{F}_{\mathbf{p}}(\mathbf{S}) = \frac{U}{V_{\mathbf{h}}} \sum_{n=1}^{N} A_{\mathbf{p},n} \exp[i(\gamma + 2\pi \mathbf{p} \cdot \mathbf{S}_n)]$$

or

$$\mathbf{F}_{\mathbf{p}}(\mathbf{S}) = \frac{U}{V_{\mathbf{h}}} \sum_{n=1}^{N} A_{\mathbf{p},n} \exp[i(\gamma_n + 2\pi \mathbf{p}_n \cdot \mathbf{S})],$$

where $\mathbf{p}_n = [\boldsymbol{C}_n^T]\mathbf{p}$ and $\mathbf{S} = \mathbf{S}_1$. Hence

$$|\mathbf{F}_{\mathbf{p}}(\mathbf{S})|^2 = \left(\frac{U}{V_{\mathbf{h}}}\right)^2 \sum_n \sum_m \big(A_{\mathbf{p},n} A_{\mathbf{p},m} \times \exp\{i[2\pi(\mathbf{p}_n - \mathbf{p}_m) \cdot \mathbf{S} + (\gamma_n - \gamma_m)]\}\big),$$

and then from (2.3.7.3)

$$T(\mathbf{S}) = \left(\frac{U}{V_{\mathbf{h}}}\right)^2 \sum_{\mathbf{p}} \sum_n \sum_m \Big(|\mathbf{F}_{\mathbf{p},\,\mathrm{obs}}|^2 A_{\mathbf{p},n} A_{\mathbf{p},m} \times \exp\{i[2\pi(\mathbf{p}_n - \mathbf{p}_m) \cdot \mathbf{S} + (\gamma_n - \gamma_m)]\}\Big), \qquad (2.3.7.4)$$

which is a Fourier summation with known coefficients $\{|\mathbf{F}_{\mathbf{p},\,\mathrm{obs}}|^2 A_{\mathbf{p},n} A_{\mathbf{p},m} \times \exp[i(\gamma_n - \gamma_m)]\}$ such that $T(\mathbf{S})$ will be a maximum at the correct molecular position.

Terms with $n = m$ in expression (2.3.7.4) can be omitted as they are independent of **S** and only contribute a constant to the value of $T(\mathbf{S})$. For terms with $n \neq m$, the indices take on special values. For instance, if the **p** cell is monoclinic with its unique axis parallel to **b** such that $\mathbf{p}_1 = (p, q, r)$ and $\mathbf{p}_2 = (\bar{p}, q, \bar{r})$, then $\mathbf{p}_1 - \mathbf{p}_2$ would be $(2p, 0, 2r)$. Hence, $T(\mathbf{S})$ would be a two-dimensional function consistent with the physical requirement that the translation component, parallel to the twofold monoclinic axis, is arbitrary.

Crowther & Blow (1967) show that if $\mathbf{F}_{\mathrm{M}}$ are the structure factors of a known molecule correctly oriented within the cell of the unknown structure at an arbitrary molecular origin, then (altering the notation very slightly from above)

$$T(\mathbf{S}) = \sum_{\mathbf{p}} |\mathbf{F}_{\mathrm{obs}}(\mathbf{p})|^2 \mathbf{F}_{\mathrm{M}}(\mathbf{p}) \mathbf{F}_{\mathrm{M}}^*(\mathbf{p}[\boldsymbol{C}]) \exp(-2\pi i \mathbf{p} \cdot \mathbf{S}),$$

where $[\boldsymbol{C}]$ is a crystallographic symmetry operator relative to which the molecular origin is to be determined. This is of the same form as (2.3.7.4) but concerns the special case where the **h** cell, into which the known molecule was placed, has the same dimensions as the **p** cell.

The translation function as defined by (2.3.7.4) is on an arbitrary scale, which makes it difficult to compare results from different calculations. Translation functions can also be defined based on the crystallographic R factor or a correlation coefficient (CC). In particular, CCs based on reflection intensities can be evaluated by Fourier methods (Navaza & Vernoslova, 1995), although it is still computationally more expensive than the evaluation of (2.3.7.4). Alternatively, the translation function can be calculated first with (2.3.7.4), and then the R factor and CC can be calculated for the resulting top solutions.

A correct solution should also produce satisfactory packing arrangements of the molecular models in the crystal. Packing functions have been derived that estimate the amount of overlap among the models (Harada *et al.*, 1981; Hendrickson & Ward, 1976; Rabinovich & Shakked, 1984; Simpson *et al.*, 2001), and such considerations can frequently limit the search volume very considerably. Alternatively, a simple enumeration of the actual close contacts among different molecules in the crystal (for example, Cα–Cα distances less than 3 Å) has also been found to be an effective way of eliminating those solutions that produce unreasonable crystal packing (Jogl *et al.*, 2001; Tong, 1993). If conformational differences are expected between the search atomic model and the actual structure, care must be taken when applying this packing check.

2.3.7.4. Position of a noncrystallographic symmetry element in a poorly defined electron-density map

If an initial set of poor phases, for example from an SIR derivative, are available and the rotation function has given the orientation of a noncrystallographic rotation axis, it is possible to search the electron-density map systematically to determine the translation axis position. The translation function must, therefore, measure the quality of superposition of the poor electron-density map on itself. Hence $\mathbf{S}_x = \mathbf{S}_{x'} = \mathbf{S}$ and the function (2.3.7.1) now becomes

$$T(\mathbf{S}) = \frac{2}{V_{\mathbf{h}}^2} \sum_{\mathbf{h}} \sum_{\mathbf{p}} |\mathbf{F}_{\mathbf{h}}||\mathbf{F}_{\mathbf{p}}| G_{\mathbf{hp}} \cos[\alpha_{\mathbf{h}} + \alpha_{\mathbf{p}} - 2\pi(\mathbf{h} + \mathbf{p}) \cdot \mathbf{S}].$$

This real-space translation function has been used successfully to determine the intermolecular dyad axis for α-chymotrypsin (Blow *et al.*, 1964) and to verify the position of immunoglobulin domains (Colman & Fehlhammer, 1976).

2.3.7.5. Locked translation function

In a translation search, an atomic model with a given orientation is moved systematically through the unit cell. In such a situation, the structure-factor equation takes on the special form (Harada *et al.*, 1981; Rae, 1977; Tong, 1993)

$$\mathbf{F}_{\mathbf{h}}^{\mathrm{c}} = \sum_n \mathbf{F}_{\mathbf{h},n} \exp(2\pi i \mathbf{h}^T [\mathbf{T}_n] \mathbf{S}),$$

where **S** is the translation vector and the summation goes over the crystallographic symmetry operators. $\mathbf{F}_{\mathbf{h},n}$ is the structure factor calculated based only on the nth symmetry-related molecule,

$$\mathbf{F}_{\mathbf{h},n} = \sum_j f_j \exp\{2\pi i \mathbf{h}^T ([\mathbf{T}_n] \mathbf{x}_j^0 + \mathbf{t}_n)\},$$

where $\mathbf{x}_j^0$ represents the atomic position of the model at the reference position and the summation goes over all the atoms.

Noting equation (2.3.7.3), the translation function is given by

$$T(\mathbf{S}) = \sum_{\mathbf{h}} \sum_n |\mathbf{F}_{\mathbf{h}}^{\mathrm{o}}|^2 |\mathbf{F}_{\mathbf{h},n}|^2 + \sum_{\mathbf{h}} \sum_n \sum_{m \neq n} |\mathbf{F}_{\mathbf{h}}^{\mathrm{o}}|^2 |\mathbf{F}_{\mathbf{h},n} \mathbf{F}_{\mathbf{h},m}^*| \exp\{-2\pi i \mathbf{h}^T ([\mathbf{T}_m] - [\mathbf{T}_n]) \mathbf{S}\}, \qquad (2.3.7.5)$$

where the second term is the ordinary translation function, analogous to (2.3.7.4). The first term of (2.3.7.5) depends on the orientation of the model. Maximization of this term, or its correlation coefficient equivalent, is the basis behind the Patterson-correlation refinement (Brünger, 1990; Tong, 1996*b*) and the direct rotation function (DeLano & Brünger, 1995). It is

also related to the intensity-based domain refinement (Yeates & Rini, 1990).

In the presence of noncrystallographic symmetry, the locked self rotation function can be used to define the orientation of the noncrystallographic symmetry point group in the crystal. If an atomic model is available for the monomer but not for the entire oligomer, the locked cross rotation function can be used to determine the orientation of this monomer in the oligomer. The locked translation function can then be used to determine the position of this monomer relative to the centre of the noncrystallographic symmetry point group (Tong, 1996*b*, 2001*a*), which will produce a model for the entire oligomer. The centre of this oligomer in the crystal can be defined by a simple translation search.

With the knowledge of the orientation of one monomer of the oligomer, the first term of (2.3.7.5) is dependent on the position of this monomer relative to the centre of the noncrystallographic symmetry oligomer (Tong, 1996*b*). The atomic positions of the entire noncrystallographic symmetry oligomer in the standard orientation are given by

$$\mathbf{X}_{n,j} = [\mathbf{I}_n]\big([\mathbf{F}]\mathbf{X}_j^0 + \mathbf{V}_0\big),$$

where $\mathbf{X}_j^0$ are the atomic positions of the monomer model, centred at (0, 0, 0); [**F**] is the orientation of this model in the oligomer in the standard orientation; $\mathbf{V}_0$ is the position of this monomer relative to the centre of the oligomer; and $[\mathbf{I}_n]$ is the nth noncrystallographic symmetry rotation matrix in the standard orientation. The atomic positions of the noncrystallographic symmetry oligomer in the crystal unit cell, centred at the origin, are given by

$$\mathbf{x}_{n,j} = [\boldsymbol{\alpha}][\mathbf{E}]\mathbf{X}_{n,j} = [\boldsymbol{\alpha}][\mathbf{E}][\mathbf{I}_n]\big([\mathbf{F}]\mathbf{X}_j^0 + \mathbf{V}_0\big),$$

where [**E**] is the orientation of the noncrystallographic symmetry in the crystal unit cell and [$\boldsymbol{\alpha}$] is the deorthogonalization matrix.

By incorporating the calculated structure factors based on this noncrystallographic symmetry oligomer into the first term of (2.3.7.5), the locked translation function is given by

$$\begin{aligned} T_{\rm L}(\mathbf{V}_0) &= \sum_{\mathbf{h}} |\mathbf{F}_{\mathbf{h}}^{\rm o}|^2 |\mathbf{F}_{\mathbf{h}}|^2 \\ &= \sum_{\mathbf{h}} \sum_{n} \sum_{m \neq n} |\mathbf{F}_{\mathbf{h}}^{\rm o}|^2 \mathbf{F}_{\mathbf{h},n} \mathbf{F}_{\mathbf{h},m}^{*} \exp\{-2\pi i \mathbf{h}([\boldsymbol{\theta}_m] - [\boldsymbol{\theta}_n])\mathbf{V}_0\}, \end{aligned} \tag{2.3.7.6}$$

where $[\boldsymbol{\theta}_n] = [\boldsymbol{\alpha}][\mathbf{E}][\mathbf{I}_n]$ and $\mathbf{F}_{\mathbf{h},n_o} = \sum_j f_j \exp(2\pi i \mathbf{h}[\boldsymbol{\theta}_n][\mathbf{F}]\mathbf{X}_j^0)$. A constant term $\sum_{\mathbf{h}} \sum_n |\mathbf{F}_{\mathbf{h}}^{\rm o}|^2 |\mathbf{F}_{\mathbf{h},n}|^2$ has been omitted from this equation.

Conceptually, the locked translation function is based on the overlap of intermolecular vectors within the noncrystallographic symmetry oligomer and the observed Patterson map (Tong, 1996*b*). The equation for the locked translation function, (2.3.7.6), bears remarkable resemblance to that for the ordinary Patterson-correlation translation function, (2.3.7.5), with the interchange of the crystallographic ([$\mathbf{T}_n$]) and noncrystallographic symmetry ([$\boldsymbol{\theta}_n$]) parameters.

2.3.7.6. Computer programs for rotation and translation function calculations

Several programs are currently in popular use for the calculation of rotation and translation functions. These include *AMoRe* (Navaza, 1994, 2001*a*), *BEAST* (Read, 2001*b*), *CCP4* (Collaborative Computational Project, Number 4, 1994), *CNS* (Brünger *et al.*, 1998), *COMO* (Jogl *et al.*, 2001), *EPMR* (Kissinger *et al.*, 1999), *GLRF* (part of the *Replace* package) (Tong, 1993, 2001*a*; Tong & Rossmann, 1990, 1997), *Molrep* (Vagin & Teplyakov, 2000) and *Phaser* (Storoni *et al.*, 2004).

The correct placement of an atomic model in a crystal unit cell is generally a six-dimensional problem, with three degrees of rotational freedom and three degrees of translational freedom. Systematic examination of all six degrees of freedom at the same time is computationally expensive and cannot be used routinely (Fujinaga & Read, 1987; Rabinovich & Shakked, 1984; Sheriff *et al.*, 1999). On the other hand, directed sampling of the six degrees of freedom, driven by a stochastic or genetic algorithm (Chang & Lewis, 1997; Glykos & Kokkinidis, 2000; Kissinger *et al.*, 1999), has been successful in solving structures.

Traditionally, the calculations are divided into a rotational component (the rotation function) and a translational component (the translation function). Only a few rotation angles (for example the top few peaks of the rotation function) are manually passed to the translation function for examination (Fitzgerald, 1988). With the power of modern computers, it is now possible to perform limited six-dimensional searches, with the sampling of the rotational degrees of freedom guided by the rotation function. For example, the top peaks of the rotation function (Navaza, 1994) and their neighbours (Urzhumtsev & Podjarny, 1995) can be automatically examined by the translation function. A more general approach is to examine all rotation-function grid points with values greater than a certain threshold (Tong, 1996*a*). Such combined molecular replacement protocols have been found to be very powerful in solving new structures.

2.3.8. Molecular replacement

2.3.8.1. Using a known molecular fragment

The most straightforward application of the molecular replacement method occurs when the orientation and position of a known molecular fragment in an unknown cell have been previously determined. The simple procedure is to apply the rotation and translation operations to the known fragment. This will place it into one 'standard' asymmetric unit of the unknown cell. Then the crystal operators (assuming no further noncrystallographic operators are present in the unknown cell) are applied to generate the complete unit cell of the unknown structure. Structure factors can then be calculated from the rotated and translated known molecule into the unknown cell. The resultant model can be refined in numerous ways.

More generally, consider a molecule placed in any crystal cell (**h**), within which coordinate positions shall be designated by **x**. Let the corresponding structure factors be $\mathbf{F}_{\mathbf{h}}$. It is then possible to compute the structure factors $\mathbf{F}_{\mathbf{p}}$ for another cell (**p**) into which the same molecule has been placed N times related by the crystallographic symmetry operators $[\boldsymbol{C}_1], \mathbf{d}_1; [\boldsymbol{C}_2], \mathbf{d}_2; \ldots; [\boldsymbol{C}_N], \mathbf{d}_N$. Let the electron density at a point $\mathbf{y}_1$ in the first crystallographic asymmetric unit be spatially related to the point $\mathbf{y}_n$ in the nth asymmetric unit of the **p** crystal such that

$$\rho(\mathbf{y}_n) = \rho(\mathbf{y}_1), \tag{2.3.8.1}$$

where

$$\mathbf{y}_n = [\boldsymbol{C}_n]\mathbf{y}_1 + \mathbf{d}_n. \tag{2.3.8.2}$$

From the definition of a structure factor,

$$\mathbf{F}_{\mathbf{p}} = \sum_{n=1}^{N} \int_U \rho(\mathbf{y}_n) \exp(2\pi i \mathbf{p} \cdot \mathbf{y}_n)\, {\rm d}\mathbf{y}_n, \tag{2.3.8.3}$$

where the integral is taken over the volume U of one molecule. But since each molecule is identical as expressed in equation (2.3.8.1) and since (2.3.8.2) can be substituted in equation (2.3.8.3), we have

$$\mathbf{F}_\mathbf{p} = \sum_{n=1}^{N} \int_U \rho(\mathbf{y}_1) \exp[2\pi i \mathbf{p} \cdot ([\boldsymbol{C}_n]\mathbf{y}_1 + \mathbf{d}_n)] \, \mathrm{d}\mathbf{y}_1. \quad (2.3.8.4)$$

Now let the molecule in the **h** crystal be related to the molecule in the first asymmetric unit of the **p** crystal by the noncrystallographic symmetry operation

$$\mathbf{x} = [\boldsymbol{C}]\mathbf{y} + \mathbf{d}, \quad (2.3.8.5)$$

which implies

$$\rho(\mathbf{x}) = \rho(\mathbf{y}_1) = \rho(\mathbf{y}_2) = \ldots. \quad (2.3.8.6)$$

Furthermore, in the **h** cell

$$\rho(\mathbf{x}) = \frac{1}{V_\mathbf{h}} \sum_\mathbf{h} \mathbf{F}_\mathbf{h} \exp(-2\pi i \mathbf{h} \cdot \mathbf{x}), \quad (2.3.8.7)$$

and thus, by combining with (2.3.8.5), (2.3.8.6) and (2.3.8.7),

$$\rho(\mathbf{y}_1) = \frac{1}{V_\mathbf{h}} \sum_\mathbf{h} \mathbf{F}_\mathbf{h} \exp[-2\pi i(\mathbf{h}[\boldsymbol{C}] \cdot \mathbf{y}_1 + \mathbf{h} \cdot \mathbf{d})]. \quad (2.3.8.8)$$

Now using (2.3.8.4) and (2.3.8.8) it can be shown that

$$\mathbf{F}_\mathbf{p} = \frac{U}{V_\mathbf{h}} \sum_\mathbf{h} \mathbf{F}_\mathbf{h} \sum_{n=1}^{N} G_{\mathbf{hp}_n} \exp[2\pi i(\mathbf{p} \cdot \mathbf{S}_n - \mathbf{h} \cdot \mathbf{S})], \quad (2.3.8.9)$$

where

$$UG_{\mathbf{hp}_n} = \int_U \exp[2\pi i(p[\boldsymbol{C}_n] - \mathbf{h}[\boldsymbol{C}]) \cdot \mathbf{u}] \, \mathrm{d}\mathbf{u}. \quad (2.3.8.10)$$

S is a chosen molecular origin in the **h** crystal and $\mathbf{S}_n$ is the corresponding molecular position in the nth asymmetric unit of the **p** crystal.

2.3.8.2. Using noncrystallographic symmetry for phase improvement

The use of noncrystallographic symmetry for phase determination was proposed by Rossmann & Blow (1962, 1963) and subsequently explored by Crowther (1967, 1969) and Main & Rossmann (1966). These methods were developed in reciprocal space and were primarily concerned with *ab initio* phase determination. Real-space averaging of electron density between noncrystallographically related molecules was used in the structure determination of deoxyhaemoglobin (Muirhead *et al*., 1967) and of α-chymotrypsin (Matthews *et al*., 1967). The improvement derived from the averaging between the two noncrystallographic units was, however, not clear in either case. The first obviously successful application was in the structure determination of lobster glyceraldehyde-3-phosphate dehydrogenase (Buehner *et al*., 1974; Argos *et al*., 1975), where the tetrameric molecule of symmetry 222 occupied one crystallographic asymmetric unit. The improvement in the essentially SIR electron-density map was considerable and the results changed from uninterpretable to interpretable. The uniqueness and validity of the solution lay in the obvious chemical correctness of the polypeptide fold and its agreement with known amino-acid-sequence data. In contrast to the earlier reciprocal-space methods, noncrystallographic symmetry was used as a method to improve poor phases rather than to determine phases *ab initio*.

Table 2.3.8.1. *Molecular replacement: phase refinement as an iterative process*

(*A*)	$\mathbf{F}_\mathrm{obs}, \alpha'_n, m'_n \rightarrow \rho_n$
(*B*)	$\rho_n \rightarrow \rho_n$(modified) (i) Use of noncrystallographic symmetry operators (ii) Definition of envelope limiting volume within which noncrystallographic symmetry is valid (iii) Adjustment of solvent density† (iv) Use of crystallographic operators to reconstruct modified density into a complete cell
(*C*)	ρ_n(modified) $\rightarrow \mathbf{F}_{\mathrm{calc},\,n+1}; \alpha_{\mathrm{calc},\,n+1}$
(*D*)	$(\mathbf{F}_{\mathrm{calc},\,n+1}, \alpha_{\mathrm{calc},\,n+1}) + (\mathbf{F}_\mathrm{obs}, \alpha_0) \rightarrow \mathbf{F}_\mathrm{obs}, \alpha'_{n+1}, m'_{n+1}$ (i) Assessment of reliability of new phasing set α_{n+1} in relation to original phasing set $\alpha_0(w)$ (ii) Use of figures of merit m_0, m_{n+1} and reliability w to determine modified phasing set α'_{n+1}, m'_{n+1}‡ (iii) Consideration of α_{n+1} and m_{n+1} where there was no prior knowledge of (*a*) $\mathbf{F}_\mathrm{obs}$ (*e.g.* very low order reflections or uncollected data) (*b*) α_0 (*e.g.* no isomorphous information or phase extension)
(*E*)	Return to step (*A*) with α'_{n+1}, m'_{n+1} and a possibly augmented set of $\mathbf{F}_\mathrm{obs}$.

† Wang (1985); Bhat & Blow (1982); Collins (1975); Schevitz *et al*. (1981); Hoppe & Gassmann (1968). ‡ Rossmann & Blow (1961); Hendrickson & Lattman (1970).

Many other applications followed rapidly, aided greatly by the versatile techniques developed by Bricogne (1976). Of particular interest is the application to the structure determination of hexokinase (Fletterick & Steitz, 1976), where the averaging occurred both between different crystal forms and within the same crystal.

The most widely used procedure for real-space averaging is the 'double sorting' technique developed by Bricogne (1976) and also by Johnson (1978). An alternative method is to maintain the complete map stored in the computer (Nordman, 1980*b*). This avoids the sorting operation, but is only possible given a very large computer or a low-resolution map containing relatively few grid points.

Bricogne's double sorting technique involves generating real-space non-integral points (D_i) which are related to integral grid points (I_i) in the cell asymmetric unit by the noncrystallographic symmetry operators. The elements of the set D_i are then brought back to their equivalent points in the cell asymmetric unit (D'_i) and sorted by their proximity to two adjacent real-space sections. The set I'_i, calculated on a finer grid than I_i and stored in the computer memory two sections at a time, is then used for linear interpolation to determine the density values at D'_i which are successively stored and summed in the related array I_i. A count is kept of the number of densities received at each I_i, resulting in a final averaged aggregate, when all real-space sections have been utilized. The density to be assigned outside the molecular envelope (defined with respect to the set I_i) is determined by averaging the density of all unused points in I_i. The grid interval for the set I'_i should be about one-sixth of the resolution to avoid serious errors from interpolation (Bricogne, 1976). The grid point separation in the set I_i need only be sufficient for representation of electron density, or about one-third of the resolution.

Molecular replacement in real space consists of the following steps (Table 2.3.8.1): (*a*) calculation of electron density based on a starting phase set and observed amplitudes; (*b*) averaging of this density among the noncrystallographic asymmetric units or molecular copies in several crystal forms, a process which defines

a molecular envelope as the averaging is only valid within the range of the noncrystallographic symmetry; (*c*) reconstructing the unit cell based on averaged density in every noncrystallographic asymmetric unit; (*d*) calculating structure factors from the reconstructed cell; (*e*) combining the new phases with others to obtain a weighted best-phase set; and (*f*) returning to step (*a*) at the previous or an extended resolution. Decisions made in steps (*b*) and (*e*) determine the rate of convergence (see Table 2.3.8.1) to a solution (Arnold *et al*., 1987).

The power of the molecular replacement procedure for either phase improvement or phase extension depends on the number of noncrystallographic asymmetric units, the size of the excluded volume expressed in terms of the ratio $(V - UN)/V$ and the magnitude of the measurement error on the structure amplitudes. Crowther (1967, 1969) and Bricogne (1974) have investigated the dependence on the number of noncrystallographic asymmetric units and conclude that three or more copies are sufficient to ensure convergence of an iterative phase improvement procedure in the absence of errors on the structure amplitudes. As with the analogous case of isomorphous replacement in which three data sets ensure reasonable phase determination, additional copies will enhance the power of the method, although their usefulness is subject to the law of diminishing returns. Another example of this principle is the sign determination of the *h*0*l* reflections of horse haemoglobin (Perutz, 1954) in which seven shrinkage stages constituted the sampling of the transform of a single copy.

In an analysis of how phasing errors propagate into errors in calculations of electron density, Arnold & Rossmann (1986) concluded that the 'power' of phase determination could be related to the noncrystallographic redundancy, *N*, the ratio of the molecular envelope volume, *U*, to the unit cell volume, *V*, the fractional error of the structure-factor amplitudes, *R* and the fractional completeness of the data, *f*, by (Arnold & Rossmann, 1986)

$$P = \frac{(Nf)^{1/2}}{RU/V}. \qquad (2.3.8.11)$$

This semiquantitative result makes intuitive sense in that the noncrystallographic redundancy and solvent content terms can be directly related to over-sampling of the molecular transform in reciprocal space, and, thus, are analogous in providing phasing information. The phasing power of solvent flattening/density modification was further analysed and shown to lead to Sayre's equations (Sayre, 1952) at a limit where the molecular envelope is sufficiently detailed and shrunken to cover sharpened and separated atoms (Arnold & Rossmann, 1986). This result suggests that more detailed definitions of molecular envelopes than are traditionally used could be advantageous for phase improvement and extension procedures.

Procedures for real-space averaging have been used extensively with great success. The interesting work of Wilson *et al*. (1981) is noteworthy for the continuous adjustment of molecular envelope with increased map definition. Furthermore, the analysis of complete virus structures has only been possible as a consequence of this technique (Bloomer *et al*., 1978; Harrison *et al*., 1978; Abad-Zapatero *et al*., 1980; Liljas *et al*., 1982). Although the procedure has been used primarily for phase improvement, apparently successful attempts have been made at phase extension (Nordman, 1980*b*; Gaykema *et al*., 1984; Rossmann *et al*., 1985). *Ab initio* phasing of glyceraldehyde-3-phosphate dehydrogenase (Argos *et al*., 1975) was successfully attempted by initially filling the known envelope with uniform density to determine the phases of the innermost reflections and then gradually extending phases to 6.3 Å resolution. Johnson *et al*. (1976) used the same procedure to determine the structure of southern bean mosaic virus to 22.5 Å resolution. Particularly impressive was the work on polyoma virus (Rayment *et al*., 1982; Rayment, 1983; Rayment *et al*., 1983) where crude initial models led to an entirely unexpected breakdown of the Caspar & Klug (1962) concept of quasi-symmetry. *Ab initio* phasing has also been used by combining the electron-diffraction projection data of two different crystal forms of bacterial rhodopsin (Rossmann & Henderson, 1982).

2.3.8.3. Update on noncrystallographic averaging and density-modification methods

Since this article was originally written, molecular replacement has been subject of a number of reviews (Rossmann, 1990), including a historical background of the subject (Rossmann, 2001). A series of chapters pertaining to molecular replacement have been published in *IT* Volume F (Rossmann & Arnold, 2001*a*), reviewing noncrystallographic symmetry (Chapter 13.1; Blow, 2001), rotation (Chapter 13.2; Navaza, 2001*b*) and translation (Chapter 13.3; Tong, 2001*b*) functions, and noncrystallographic symmetry averaging for phase improvement and extension (Chapter 13.4; Rossmann & Arnold, 2001*b*). Chapters on phase improvement by density modification (Chapter 15.1; Zhang *et al.*, 2001), optimal weighting of Fourier terms in map calculations (Chapter 15.2; Read, 2001*a*) and refinement calculations incorporating bulk solvent correction (Chapter 18.4; Dauter *et al.*, 2001) are also recommended reading.

There has been remarkable progress in the general area of density modification, involving improvement of real-space methods for averaging and reconstruction, and treatment of solvent for iterative phase improvement and refinement calculations. The use of real-space averaging between noncrystallographically related electron density within the crystallographic asymmetric unit has become an accepted mode of extending phase information to higher resolution, particularly for complex structures such as viruses [Acharya *et al.*, 1989; Arnold & Rossmann, 1988; Gaykema *et al.*, 1986; Hogle *et al.*, 1985; Luo *et al.*, 1989; Rossmann & Arnold, 2001*b* (*IT* F Chapter 13.4); Rossmann *et al.*, 1985, 1992]. *Ab initio* phase determination based on noncrystallographic redundancy has become fairly common (Chapman *et al.*, 1992; Lunin *et al.*, 2000; Miller *et al.*, 2001; Rossmann, 1990; Tsao *et al.*, 1992). General programs in common use for noncrystallographic symmetry averaging include *BUSTER-TNT* [Blanc *et al.*, 2004; Roversi *et al.*, 2000; Tronrud & Ten Eyck, 2001 (*IT* F Section 25.2.4)], *CNS* [Brünger *et al.*, 1998; Brunger, Adams, DeLano *et al.*, 2001 (*IT* F Section 25.2.3)], *DM*/*DMMULTI* [Cowtan & Main, 1993; Cowtan *et al.*, 2001 (*IT* F Section 25.2.2); Schuller, 1996; Zhang, 1993], *PHASES* [Furey, 2001 (*IT* F Section 25.2.1); Furey & Swaminathan, 1997], *RAVE*/*MAVE* (Jones, 1992; Kleywegt, 1996) and *SOLVE*/*RESOLVE* [Terwilliger, 2002*b*, 2003*c*; Terwilliger & Berendzen, 2001 (*IT* F Section 14.2.2)].

Solvent flattening has been formulated in reciprocal space for greater computational efficiency (Leslie, 1987; Terwilliger, 1999) and solvent 'flipping' is a powerful extension of solvent density modification (Abrahams, 1997; Abrahams & Leslie, 1996). Bulk-solvent corrections are now commonly used in crystallographic refinement, allowing for better modelling and phase determination of low-resolution data [Brünger *et al.*, 1998; Dauter *et al.*, 2001 (*IT* F Chapter 18.4)]. The problem of phase error estimation and analysis and bias removal has been treated extensively (Cowtan, 1999; Cowtan & Main, 1996), including extension of methods to include maximum-likelihood functions and iterative bias removal procedures [Brunger, Adams & Rice, 2001 (*IT* F Chapter 18.2); Hunt & Deisenhofer, 2003; Lamzin *et al.*, 2001 (*IT* F Section 25.2.5); Perrakis *et al.*, 1997; Terwilliger, 2004]. Histogram matching [Cowtan & Main, 1993; Lunin, 1993; Nieh & Zhang, 1999; Refaat *et al.*, 1996; Zhang, 1993; Zhang *et al.*, 2001 (*IT* F Chapter 15.1)] and skeletonization [Baker *et al.*, 1993;

Zhang *et al.*, 2001 (*IT* F Chapter 15.1)], and structural fragment matching procedures (Terwilliger, 2003*a*) have been added to the arsenal of density-modification methods. Automated mask and molecular-envelope definition has helped to remove the tedium and increase the efficiency and quality of density-modification and symmetry-averaging procedures. Noncrystallographic symmetry averaging among different crystal forms (Perutz, 1954) has become increasingly common, and exploitation of the unit-cell variation among flash-cooled and noncooled forms of the same crystal is a broadly applicable method for phase determination (Das *et al.*, 1996; Ding *et al.*, 1995); soaking crystals in a series of different solvents and buffers can produce an analogous effect (Ren *et al.*, 1995; Tong *et al.*, 1997). Phases from noncrystallographic symmetry averaging and other 'experimental' sources have been incorporated into crystallographic refinement procedures using a number of formalisms (Arnold & Rossmann, 1988; Rees & Lewis, 1983) including maximum likelihood (Pannu *et al.*, 1998).

2.3.8.4. Equivalence of real- and reciprocal-space molecular replacement

Let us proceed in reciprocal space doing exactly the same as is done in real-space averaging. Thus

$$\rho_{\rm AV}(\mathbf{x}) = \frac{1}{N}\sum_{n=1}^{N}\rho(\mathbf{x}_n),$$

where

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

Therefore,

$$\rho_{\rm AV}(\mathbf{x}) = \frac{1}{N}\sum_{N}\frac{1}{V}\left[\sum_{\mathbf{h}}\mathbf{F}_{\mathbf{h}}\exp(2\pi i\mathbf{h}\cdot\mathbf{x}_n)\right].$$

The next step is to perform the back-transform of the averaged electron density. Hence,

$$\mathbf{F}_{\mathbf{p}} = \int_U \rho_{\rm AV}(\mathbf{x})\exp(-2\pi i\mathbf{p}\cdot\mathbf{x})\,{\rm d}\mathbf{x},$$

where U is the volume within the averaged part of the cell. Hence, substituting for $\rho_{\rm AV}$,

$$\mathbf{F}_{\mathbf{p}} = \int_U\left[\frac{1}{NV}\sum_{N}\sum_{\mathbf{h}}\mathbf{F}_{\mathbf{h}}\exp(2\pi i\mathbf{h}\cdot\mathbf{x}_n)\right]\exp(-2\pi i\mathbf{p}\cdot\mathbf{x})\,{\rm d}\mathbf{x},$$

which is readily simplified to

$$\mathbf{F}_{\mathbf{p}} = \frac{U}{NV}\sum_{\mathbf{h}}\mathbf{F}_{\mathbf{h}}\sum_{N}G_{\mathbf{hp}n}\exp(2\pi i\mathbf{h}\cdot\mathbf{d}_n).$$

Setting

$$\mathbf{B}_{\mathbf{hp}} = \frac{U}{NV}\sum_{N}G_{\mathbf{hp}n}\exp(2\pi i\mathbf{h}\cdot\mathbf{d}_n),$$

the molecular replacement equations can be written as

$$\mathbf{F}_{\mathbf{p}} = \sum_{\mathbf{h}}\mathbf{B}_{\mathbf{hp}}\mathbf{F}_{\mathbf{h}} \qquad (2.3.8.12)$$

(Main & Rossmann, 1966), or in matrix form

$$\mathbf{F} = [\boldsymbol{B}]\mathbf{F},$$

which is the form of the equations used by Main (1967) and by Crowther (1967). Colman (1974) arrived at the same conclusions by an application of Shannon's sampling theorem. It should be noted that the elements of $[\boldsymbol{B}]$ are dependent only on knowledge of the noncrystallographic symmetry and the volume within which it is valid. Substitution of approximate phases into the right-hand side of (2.3.8.12) produces a set of calculated structure factors exactly analogous to those produced by back-transforming the averaged electron density in real space. The new phases can then be used in a renewed cycle of molecular replacement. The reciprocal-space molecular replacement procedure has been implemented and tested in a computer program (Tong & Rossmann, 1995).

Computationally, it has been found more convenient and faster to work in real space. This may, however, change with the advent of vector processing in 'supercomputers'. Obtaining improved phases by substitution of current phases on the right-hand side of the molecular replacement equations (2.3.8.1) seems less cumbersome than the repeated forward and backward Fourier transformation, intermediate sorting, and averaging required in the real-space procedure.

2.3.9. Conclusions

Complete interpretation of Patterson maps is no longer used frequently in structure analysis, although most determinations of heavy-atom positions of isomorphous pairs are based on Patterson analyses. Incorporation of the Patterson concept is crucial in many sophisticated techniques essential for the solution of complex problems, particularly in the application to biological macromolecular structures. Patterson techniques provide important physical insights in a link between real- and reciprocal-space formulation of crystal structures and diffraction data.

This article, first written in December 1984 (by MGR and EA) and completed in January 1986, was published in the first edition of this volume 1993, and in a mildly revised form in the second edition in 2001. We are grateful for generous support of our laboratories from the National Science Foundation (to LT and MGR) and from the National Institutes of Health (LT, MGR and EA). We acknowledge the many authors whose insights, innovation and writings make up the subject matter of this review. We also acknowledge Sharon Wilder for her painstaking attention to detail in preparation of the original manuscript and an article by Argos & Rossmann (1980) as the source of some material in this article.

References

Abad-Zapatero, C., Abdel-Meguid, S. S., Johnson, J. E., Leslie, A. G. W., Rayment, I., Rossmann, M. G., Suck, D. & Tsukihara, T. (1980). *Structure of southern bean mosaic virus at 2.8 Å resolution. Nature (London)*, **286**, 33–39.

Abrahams, J. P. (1997). *Bias reduction in phase refinement by modified interference functions: introducing the γ correction. Acta Cryst.* D**53**, 371–376.

Abrahams, J. P. & Leslie, A. G. W. (1996). *Methods used in the structure determination of bovine mitochondrial F1 ATPase. Acta Cryst.* D**52**, 30–42.

Acharya, R., Fry, E., Stuart, D., Fox, G., Rowlands, D. & Brown, F. (1989). *The three-dimensional structure of foot-and-mouth disease virus at 2.9 Å resolution. Nature (London)*, **337**, 709–716.

Adams, M. J., Blundell, T. L., Dodson, E. J., Dodson, G. G., Vijayan, M., Baker, E. N., Harding, M. M., Hodgkin, D. C., Rimmer, B. & Sheat, S. (1969). *Structure of rhombohedral 2 zinc insulin crystals. Nature (London)*, **224**, 491–495.

Åkervall, K., Strandberg, B., Rossmann, M. G., Bengtsson, U., Fridborg, K., Johannisen, H., Kannan, K. K., Lövgren, S., Petef, G., Öberg, B., Eaker, D., Hjertén, S., Rydén, L. & Moring, I. (1972). *X-ray diffraction studies of the structure of satellite tobacco necrosis virus. Cold Spring Harbor Symp. Quant. Biol.* **36**, 469–488.

Argos, P., Ford, G. C. & Rossmann, M. G. (1975). *An application of the molecular replacement technique in direct space to a known protein structure. Acta Cryst.* A**31**, 499–506.

Argos, P. & Rossmann, M. G. (1974). *Determining heavy-atom positions using non-crystallographic symmetry. Acta Cryst.* A**30**, 672–677.

Argos, P. & Rossmann, M. G. (1976). *A method to determine heavy-atom positions for virus structures. Acta Cryst.* B**32**, 2975–2979.

Argos, P. & Rossmann, M. G. (1980). *Molecular replacement methods.* In *Theory and Practice of Direct Methods in Crystallography*, edited by M. F. C. Ladd & R. A. Palmer, pp. 361–417. New York: Plenum.

Arnold, E., Erickson, J. W., Fout, G. S., Frankenberger, E. A., Hecht, H. J., Luo, M., Rossmann, M. G. & Rueckert, R. R. (1984). *Virion orientation in cubic crystals of the human common cold virus HRV14. J. Mol. Biol.* **177**, 417–430.

Arnold, E. & Rossmann, M. G. (1986). *Effect of errors, redundancy, and solvent content in the molecular replacement procedure for the structure determination of biological macromolecules. Proc. Natl Acad. Sci. USA*, **83**, 5489–5493.

Arnold, E. & Rossmann, M. G. (1988). *The use of molecular-replacement phases for the refinement of the human rhinovirus 14 structure. Acta Cryst.* A**44**, 270–283.

Arnold, E., Vriend, G., Luo, M., Griffith, J. P., Kamer, G., Erickson, J. W., Johnson, J. E. & Rossmann, M. G. (1987). *The structure determination of a common cold virus, human rhinovirus 14. Acta Cryst.* A**43**, 346–361.

Baker, D., Bystroff, C., Fletterick, R. J. & Agard, D. A. (1993). *PRISM: topologically constrained phase refinement for macromolecular crystallography. Acta Cryst.* D**49**, 429–439.

Beevers, C. A. & Robertson, J. M. (1950). *Interpretation of the Patterson synthesis. Acta Cryst.* **3**, 164.

Beurskens, P. T. (1981). *A statistical interpretation of rotation and translation functions in reciprocal space. Acta Cryst.* A**37**, 426–430.

Bhat, T. N. & Blow, D. M. (1982). *A density-modification method for the improvement of poorly resolved protein electron-density maps. Acta Cryst.* A**38**, 21–29.

Bijvoet, J. M. (1954). *Structure of optically active compounds in the solid state. Nature (London)*, **173**, 888–891.

Bijvoet, J. M., Peerdeman, A. F. & van Bommel, A. J. (1951). *Determination of the absolute configuration of optically active compounds by means of X-rays. Nature (London)*, **168**, 271–272.

Blanc, E., Roversi, P., Vonrhein, C., Flensburg, C., Lea, S. M. & Bricogne, G. (2004). *Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT. Acta Cryst.* D**60**, 2210–2221.

Bloomer, A. C., Champness, J. N., Bricogne, G., Staden, R. & Klug, A. (1978). *Protein disk of tobacco mosaic virus at 2.8 Å resolution showing the interactions within and between subunits. Nature (London)*, **276**, 362–368.

Blow, D. M. (1958). *The structure of haemoglobin. VII. Determination of phase angles in the noncentrosymmetric [100] zone. Proc. R. Soc. London Ser. A*, **247**, 302–336.

Blow, D. M. (2001). *Noncrystallographic symmetry*. In *International Tables for Crystallography*, Vol. F, *Crystallography of Biological Macromolecules*, edited by M. G. Rossmann & E. Arnold, ch. 13.1. Dordrecht: Kluwer Academic Publishers.

Blow, D. M. & Crick, F. H. C. (1959). *The treatment of errors in the isomorphous replacement method. Acta Cryst.* **12**, 794–802.

Blow, D. M. & Rossmann, M. G. (1961). *The single isomorphous replacement method. Acta Cryst.* **14**, 1195–1202.

Blow, D. M., Rossmann, M. G. & Jeffery, B. A. (1964). *The arrangement of α-chymotrypsin molecules in the monoclinic crystal form. J. Mol. Biol.* **8**, 65–78.

Bluhm, M. M., Bodo, G., Dintzis, H. M. & Kendrew, J. C. (1958). *The crystal structure of myoglobin. IV. A Fourier projection of sperm-whale myoglobin by the method of isomorphous replacement. Proc. R. Soc. London Ser. A*, **246**, 369–389.

Blundell, T. L. & Johnson, L. N. (1976). *Protein Crystallography.* New York: Academic Press.

Bodo, G., Dintzis, H. M., Kendrew, J. C. & Wyckoff, H. W. (1959). *The crystal structure of myoglobin. V. A low-resolution three-dimensional Fourier synthesis of sperm-whale myoglobin crystals. Proc. R. Soc. London Ser. A*, **253**, 70–102.

Bragg, W. L. (1958). *The determination of the coordinates of heavy atoms in protein crystals. Acta Cryst.* **11**, 70–75.

Bragg, W. L. & Perutz, M. F. (1954). *The structure of haemoglobin. VI. Fourier projections on the 010 plane. Proc. R. Soc. London Ser. A*, **225**, 315–329.

Braun, P. B., Hornstra, J. & Leenhouts, J. I. (1969). *Automated crystal-structure determination by Patterson search using a known part of the molecule. Philips Res. Rep.* **24**, 85–118.

Bricogne, G. (1974). *Geometric sources of redundancy in intensity data and their use for phase determination. Acta Cryst.* A**30**, 395–405.

Bricogne, G. (1976). *Methods and programs for the direct space exploitation of geometric redundancies. Acta Cryst.* A**32**, 832–847.

Brünger, A. T. (1990). *Extension of molecular replacement: a new search strategy based on Patterson correlation refinement. Acta Cryst.* A**46**, 46–57.

Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Crystallography & NMR System: a new software suite for macromolecular structure determination. Acta Cryst.* D**54**, 905–921.

Brunger, A. T., Adams, P. D., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Pannu, N. S., Read, R. J., Rice, L. M. & Simonson, T. (2001). *The structure-determination language of the Crystallography & NMR System*. In *International Tables for Crystallography*, Vol. F, *Crystallography of Biological Macromolecules*, edited by M. G. Rossmann & E. Arnold, Section 25.2.3. Dordrecht: Kluwer Academic Publishers.

Brunger, A. T., Adams, P. D. & Rice, L. M. (2001). *Enhanced macromolecular refinement by simulated annealing*. In *International Tables for Crystallography*, Vol. F, *Crystallography of Biological Macromolecules*, edited by M. G. Rossmann & E. Arnold, ch. 18.2. Dordrecht: Kluwer Academic Publishers.

Buehner, M., Ford, G. C., Moras, D., Olsen, K. W. & Rossmann, M. G. (1974). *Structure determination of crystalline lobster* D*-glyceraldehyde-3-phosphate dehydrogenase. J. Mol. Biol.* **82**, 563–585.

Buerger, M. J. (1946). *The interpretation of Harker syntheses. J. Appl. Phys.* **17**, 579–595.

Buerger, M. J. (1950*a*). *Some new functions of interest in X-ray crystallography. Proc. Natl Acad. Sci. USA*, **36**, 376–382.

Buerger, M. J. (1950*b*). *Limitation of electron density by the Patterson function. Proc. Natl Acad. Sci. USA*, **36**, 738–742.

Buerger, M. J. (1951). *A new approach to crystal-structure analysis. Acta Cryst.* **4**, 531–544.

Buerger, M. J. (1953*a*). *Image theory of superposed vector sets. Proc. Natl Acad. Sci. USA*, **39**, 669–673.

Buerger, M. J. (1953*b*). *Solution functions for solving superposed Patterson syntheses. Proc. Natl Acad. Sci. USA*, **39**, 674–678.

Buerger, M. J. (1953*c*). *An intersection function and its relations to the minimum function of X-ray crystallography. Proc. Natl Acad. Sci. USA*, **39**, 678–680.

Buerger, M. J. (1959). *Vector Space and its Application in Crystal-Structure Investigation.* New York: John Wiley.

Buerger, M. J. (1966). *Background for the use of image-seeking functions. Trans. Am. Crystallogr. Assoc.* **2**, 1–9.

Bullough, R. K. (1961). *On homometric sets. I. Some general theorems. Acta Cryst.* **14**, 257–269.

Bullough, R. K. (1964). *On homometric sets. II. Sets obtained by singular transformations. Acta Cryst.* **17**, 295–308.

Burdina, V. I. (1970). *Symmetry of the rotation function. Kristallografiya*, **15**, 623–630.

Burdina, V. I. (1971). *Symmetry of the rotation function. Sov. Phys. Crystallogr.* **15**, 545–550.

Burdina, V. I. (1973). *Primitive rotation regions of two Patterson syntheses. Kristallografiya*, **18**, 694–700.

Burnett, R. M. & Rossmann, M. G. (1971). *The determination of the crystal structure of trans-2,4-dihydroxy-2,4-dimethylcyclohexane-trans-*

1-acetic acid γ-lactone, $C_{10}H_{16}O_3$, using rotation and translation functions in reciprocal space. Acta Cryst. B**27**, 1378–1387.
Carlisle, C. H. & Crowfoot, D. (1945). *The crystal structure of cholesteryl iodide. Proc. R. Soc. London Ser. A*, **184**, 64–83.
Caspar, D. L. D. & Klug, A. (1962). *Physical principles in the construction of regular viruses. Cold Spring Harbor Symp. Quant. Biol.* **27**, 1–24.
Chang, G. & Lewis, M. (1997). *Molecular replacement using genetic algorithms. Acta Cryst.* D**53**, 279–289.
Chapman, M. S., Tsao, J. & Rossmann, M. G. (1992). *Ab initio phase determination for spherical viruses: parameter determination for spherical-shell models*. *Acta Cryst.* A**48**, 301–312.
Clastre, J. & Gay, R. (1950). *La détermination des structures cristallines à partir du diagramme de Patterson. Compt. Rend.* **230**, 1876–1877.
Collaborative Computational Project, Number 4 (1994). *The CCP4 suite: programs for protein crystallography*. *Acta Cryst.* D**50**, 760–763.
Collins, D. M. (1975). *Efficiency in Fourier phase refinement for protein crystal structures. Acta Cryst.* A**31**, 388–389.
Colman, P. M. (1974). *Noncrystallographic symmetry and the sampling theorem. Z. Kristallogr.* **140**, 344–349.
Colman, P. M. & Fehlhammer, H. (1976). *Appendix: the use of rotation and translation functions in the interpretation of low resolution electron density maps. J. Mol. Biol.* **100**, 278–282.
Colman, P. M., Fehlhammer, H. & Bartels, K. (1976). *Patterson search methods in protein structure determination: β-trypsin and immunoglobulin fragments.* In *Crystallographic Computing Techniques*, edited by F. R. Ahmed, K. Huml & B. Sedlacek, pp. 248–258. Copenhagen: Munksgaard.
Corfield, P. W. R. & Rosenstein, R. D. (1966). *Maximum information from the minimum function. Trans. Am. Crystallogr. Assoc.* **2**, 17–28.
Cowtan, K. (1999). *Error estimation and bias correction in phase-improvement calculations*. *Acta Cryst.* D**55**, 1555–1567.
Cowtan, K. D. & Main, P. (1993). *Improvement of macromolecular electron-density maps by the simultaneous application of real and reciprocal space constraints. Acta Cryst.* D**49**, 148–157.
Cowtan, K. D. & Main, P. (1996). *Phase combination and cross validation in iterated density-modification calculations*. *Acta Cryst.* D**52**, 43–48.
Cowtan, K. D., Zhang, K. Y. J. & Main, P. (2001). *DM/DMMULTI software for phase improvement by density modification*. In *International Tables for Crystallography*, Vol. F, *Crystallography of Biological Macromolecules*, edited by M. G. Rossmann & E. Arnold, Section 25.2.2. Dordrecht: Kluwer Academic Publishers.
Crick, F. H. C. & Magdoff, B. S. (1956). *The theory of the method of isomorphous replacement for protein crystals. I. Acta Cryst.* **9**, 901–908.
Cromer, D. T. (1974). *Dispersion corrections for X-ray atomic scattering factors.* In *International Tables for X-ray Crystallography*, Vol. IV, edited by J. A. Ibers & W. C. Hamilton, pp. 148–151. Birmingham: Kynoch Press.
Crowther, R. A. (1967). *A linear analysis of the non-crystallographic symmetry problem. Acta Cryst.* **22**, 758–764.
Crowther, R. A. (1969). *The use of non-crystallographic symmetry for phase determination. Acta Cryst.* B**25**, 2571–2580.
Crowther, R. A. (1972). *The fast rotation function.* In *The Molecular Replacement Method*, edited by M. G. Rossmann, pp. 173–178. New York: Gordon & Breach.
Crowther, R. A. & Blow, D. M. (1967). *A method of positioning a known molecule in an unknown crystal structure. Acta Cryst.* **23**, 544–548.
Cullis, A. F., Muirhead, H., Perutz, M. F., Rossmann, M. G. & North, A. C. T. (1962). *The structure of haemoglobin. IX. A three-dimensional Fourier synthesis at 5.5 Å resolution: description of the structure. Proc. R. Soc. London Ser. A*, **265**, 161–187.
Das, K., Ding, J., Hsiou, Y., Clark, A. D. Jr, Moereels, H., Koymans, L., Andries, K., Pauwels, R., Janssen, P. A. J., Boyer, P. L., Clark, P., Smith, R. H. Jr, Smith, M. B. K., Michejda, C. J., Hughes, S. H. & Arnold, E. (1996). *Crystal structures of 8-Cl and 9-Cl TIBO complexed with wild-type HIV-1 RT and 8-Cl TIBO complexed with the Tyr181Cys HIV-1 RT drug-resistant mutant. J. Mol. Biol.* **264**, 1085–1100.
Dauter, Z., Dauter, M. & Rajashankar, K. R. (2000). *Novel approach to phasing proteins: derivatization by short cryo-soaking with halides*. *Acta Cryst.* D**56**, 232–237.
Dauter, Z., Murshudov, G. N. & Wilson, K. S. (2001). *Refinement at atomic resolution*. In *International Tables for Crystallography*, Vol. F, *Crystallography of Biological Macromolecules*, edited by M. G. Rossmann & E. Arnold, ch. 18.4. Dordrecht: Kluwer Academic Publishers.
Debreczeni, J. É., Bunkóczi, G., Ma, Q., Blaser, H. & Sheldrick, G. M. (2003). *In-house measurement of the sulfur anomalous signal and its use for phasing*. *Acta Cryst.* D**59**, 688–696.
DeLano, W. L. & Brünger, A. T. (1995). *The direct rotation function: rotational Patterson correlation search applied to molecular replacement*. *Acta Cryst.* D**51**, 740–748.
Dickerson, R. E., Kendrew, J. C. & Strandberg, B. E. (1961). *The crystal structure of myoglobin: phase determination to a resolution of 2 Å by the method of isomorphous replacement. Acta Cryst.* **14**, 1188–1195.
Dickerson, R. E., Kopka, M. L., Varnum, J. C. & Weinzierl, J. E. (1967). *Bias, feedback and reliability in isomorphous phase analysis. Acta Cryst.* **23**, 511–522.
Dickerson, R. E., Weinzierl, J. E. & Palmer, R. A. (1968). *A least-squares refinement method for isomorphous replacement. Acta Cryst.* B**24**, 997–1003.
Ding, J., Das, K., Moereels, H., Koymans, L., Andries, K., Janssen, P. A. J., Hughes, S. H. & Arnold, E. (1995). *Structure of HIV-1 RT/TIBO R 86183 complex reveals similarity in the binding of diverse nonnucleoside inhibitors*. *Nature Struct. Biol.* **2**, 407–415.
Dodson, E., Harding, M. M., Hodgkin, D. C. & Rossmann, M. G. (1966). *The crystal structure of insulin. III. Evidence for a 2-fold axis in rhombohedral zinc insulin. J. Mol. Biol.* **16**, 227–241.
Egert, E. (1983). *Patterson search – an alternative to direct methods. Acta Cryst.* A**39**, 936–940.
Egert, E. & Sheldrick, G. M. (1985). *Search for a fragment of known geometry by integrated Patterson and direct methods. Acta Cryst.* A**41**, 262–268.
Eisenberg, D. (1970). *X-ray crystallography and enzyme structure.* In *The Enzymes*, edited by P. D. Boyer, Vol. I, 3rd ed., pp. 1–89. New York: Academic Press.
Fitzgerald, P. M. D. (1988). *MERLOT, an integrated package of computer programs for the determination of crystal structures by molecular replacement. J. Appl. Cryst.* **21**, 273–278.
Fletterick, R. J. & Steitz, T. A. (1976). *The combination of independent phase information obtained from separate protein structure determinations of yeast hexokinase. Acta Cryst.* A**32**, 125–132.
Fridrichsons, J. & Mathieson, A. McL. (1962). *Image-seeking. A brief study of its scope and comments on certain limitations. Acta Cryst.* **15**, 1065–1074.
Fujinaga, M. & Read, R. J. (1987). *Experiences with a new translation-function program. J. Appl. Cryst.* **20**, 517–521.
Fukuyama, K., Abdel-Meguid, S. S., Johnson, J. E. & Rossmann, M. G. (1983). *Structure of a T = 1 aggregate of alfalfa mosaic virus coat protein seen at 4.5 Å resolution. J. Mol. Biol.* **167**, 873–894.
Furey, W. (2001). *PHASES*. In *International Tables for Crystallography*, Vol. F, *Crystallography of Biological Macromolecules*, edited by M. G. Rossmann & E. Arnold, Section 25.2.1. Dordrecht: Kluwer Academic Publishers.
Furey, W. & Swaminathan, S. (1997). *PHASES-95: a program package for processing and analyzing diffraction data from macromolecules*. *Methods Enzymol.* **277**, 590–620.
Garrido, J. (1950*a*). *Sur la détermination des structures cristallines au moyen de la transformée de Patterson. Compt. Rend.* **230**, 1878–1879.
Garrido, J. (1950*b*). *Les coincidences fortuites dans la méthode des différences vectorielles. Compt. Rend.* **231**, 297–298.
Gaykema, W. P., Volbeda, A. & Hol, W. G. (1986). *Structure determination of Panulirus interruptus haemocyanin at 3.2 Å resolution. Successful phase extension by sixfold density averaging*. *J. Mol. Biol.* **187**, 255–275.
Gaykema, W. P. J., Hol, W. G. J., Vereijken, J. M., Soeter, N. M., Bak, H. J. & Beintema, J. J. (1984). *3.2 Å structure of the copper-containing, oxygen-carrying protein Panulirus interruptus haemocyanin. Nature (London)*, **309**, 23–29.
Gibbs, J. W. (1898). *Remarks regarding Michelson's letter. Nature (London)*, **59**, 200.
Glykos, N. M. & Kokkinidis, M. (2000). *A stochastic approach to molecular replacement*. *Acta Cryst.* D**56**, 169–174.
Grau, U. M., Rossmann, M. G. & Trommer, W. E. (1981). *The crystallization and structure determination of an active ternary complex of pig heart lactate dehydrogenase. Acta Cryst.* B**37**, 2019–2026.
Green, D. W., Ingram, V. M. & Perutz, M. F. (1954). *The structure of haemoglobin. IV. Sign determination by the isomorphous replacement method. Proc. R. Soc. London Ser. A*, **225**, 287–307.

Grosse-Kunstleve, R. W. & Brunger, A. T. (1999). *A highly automated heavy-atom search procedure for macromolecular structures*. *Acta Cryst.* D**55**, 1568–1577.

Hamilton, W. C. (1965). *The crystal structure of orthorhombic acetamide. Acta Cryst.* **18**, 866–870.

Harada, Y., Lifchitz, A., Berthou, J. & Jolles, P. (1981). *A translation function combining packing and diffraction information: an application to lysozyme (high-temperature form). Acta Cryst.* A**37**, 398–406.

Harker, D. (1936). *The application of the three-dimensional Patterson method and the crystal structures of proustite, Ag_3AsS_3, and pyrargyrite, Ag_3SbS_3. J. Chem. Phys.* **4**, 381–390.

Harker, D. (1956). *The determination of the phases of the structure factors of non-centrosymmetric crystals by the method of double isomorphous replacement. Acta Cryst.* **9**, 1–9.

Harrison, S. C., Olson, A. J., Schutt, C. E., Winkler, F. K. & Bricogne, G. (1978). *Tomato bushy stunt virus at 2.9 Å resolution. Nature (London)*, **276**, 368–373.

Hendrickson, W. A. (1991). *Determination of macromolecular structures from anomalous diffraction of synchrotron radiation*. *Science*, **254**, 51–58.

Hendrickson, W. A. & Lattman, E. E. (1970). *Representation of phase probability distributions for simplified combination of independent phase information. Acta Cryst.* B**26**, 136–143.

Hendrickson, W. A. & Teeter, M. M. (1981). *Structure of the hydrophobic protein crambin determined directly from the anomalous scattering of sulphur. Nature (London)*, **290**, 107–113.

Hendrickson, W. A. & Ward, K. B. (1976). *A packing function for delimiting the allowable locations of crystallized macromolecules. Acta Cryst.* A**32**, 778–780.

High, D. F. & Kraut, J. (1966). *The crystal structure of androsterone. Acta Cryst.* **21**, 88–96.

Hirshfeld, F. L. (1968). *Symmetry in the generation of trial structures. Acta Cryst.* A**24**, 301–311.

Hodgkin, D. C., Kamper, J., Lindsey, J., MacKay, M., Pickworth, J., Robertson, J. H., Shoemaker, C. B., White, J. G., Prosen, R. J. & Trueblood, K. N. (1957). *The structure of vitamin B_{12}. I. An outline of the crystallographic investigation of vitamin B_{12}. Proc. R. Soc. London Ser. A*, **242**, 228–263.

Hogle, J. M., Chow, M. & Filman, D. J. (1985). *Three-dimensional structure of poliovirus at 2.9 Å resolution. Science*, **229**, 1358–1365.

Hoppe, W. (1957*a*). *Die Faltmolekülmethode und ihre anwendung in der Röntgenographischen Konstitutionsanalyse von Biflorin ($C_{20}H_{20}O_4$). Z. Elektrochem.* **61**, 1076–1083.

Hoppe, W. (1957*b*). *Die 'Faltmolekülmethode' – eine neue Methode zur Bestimmung der Kristallstruktur bei ganz oder teilweise bekannter Molekülstruktur. Acta Cryst.* **10**, 750–751.

Hoppe, W. (1959). *Die Bestimmung genauer Schweratom-parameter in isomorphen azentrischen Kristallen. Acta Cryst.* **12**, 665–674.

Hoppe, W. (1962). *'Nahezu-Homometrische Lösungen' und Faltmolekülmethode. Z. Kristallogr.* **117**, 249–258.

Hoppe, W. & Gassmann, J. (1968). *Phase correction, a new method to solve partially known structures. Acta Cryst.* B**24**, 97–107.

Hosemann, R. & Bagchi, S. N. (1954). *On homometric structures. Acta Cryst.* **7**, 237–241.

Huber, R. (1965). *Die automatisierte Faltmolekülmethode. Acta Cryst.* **19**, 353–356.

Hughes, E. W. (1940). *The crystal structure of dicyandiamide. J. Am. Chem. Soc.* **62**, 1258–1267.

Hunt, J. F. & Deisenhofer, J. (2003). *Ping-pong cross-validation in real space: a method for increasing the phasing power of a partial model without risk of model bias*. *Acta Cryst.* D**59**, 214–224.

International Tables for Crystallography (2005). Vol. A, *Space-Group Symmetry*, edited by Th. Hahn. Heidelberg: Springer.

Jacobson, R. A., Wunderlich, J. A. & Lipscomb, W. N. (1961). *The crystal and molecular structure of cellobiose. Acta Cryst.* **14**, 598–607.

James, R. W. (1965). *The optical principles of the diffraction of X-rays.* Ithaca: Cornell University Press.

Jogl, G., Tao, X., Xu, Y. & Tong, L. (2001). *COMO: a program for combined molecular replacement*. *Acta Cryst.* D**57**, 1127–1134.

Johnson, J. E. (1978). *Appendix II. Averaging of electron density maps. Acta Cryst.* B**34**, 576–577.

Johnson, J. E., Akimoto, T., Suck, D., Rayment, I. & Rossmann, M. G. (1976). *The structure of southern bean mosaic virus at 22.5 Å resolution. Virology*, **75**, 394–400.

Johnson, J. E., Argos, P. & Rossmann, M. G. (1975). *Rotation function studies of southern bean mosaic virus at 22 Å resolution. Acta Cryst.* B**31**, 2577–2583.

Jones, T. A. (1992). *A, yaap, asap, @#*? A set of averaging programs*. In *Molecular Replacement*, edited by E. J. Dodson, S. Glover & W. Wolf, pp. 91–105. Warrington: SERC Daresbury Laboratory.

Karle, J. (1976). *Partial structures and use of the tangent formula and translation functions.* In *Crystallographic Computing Techniques*, edited by F. R. Ahmed, K. Huml & B. Sedlacek, pp. 155–164. Copenhagen: Munksgaard.

Karle, J. & Hauptman, H. (1964). *Positivity, point atoms, and Pattersons. Acta Cryst.* **17**, 392–396.

Kartha, G. (1961). *Isomorphous replacement method in non-centrosymmetric structures. Acta Cryst.* **14**, 680–686.

Kartha, G. & Parthasarathy, R. (1965). *Combination of multiple isomorphous replacement and anomalous dispersion data for protein structure determination. I. Determination of heavy-atom positions in protein derivatives. Acta Cryst.* **18**, 745–749.

Ketelaar, J. A. A. & de Vries, T. A. (1939). *The crystal structure of tetra phosphonitrile chloride, $P_4N_4Cl_8$. Recl Trav. Chim.* **58**, 1081–1099.

Kissinger, C. R., Gehlhaar, D. K. & Fogel, D. B. (1999). *Rapid automated molecular replacement by evolutionary search*. *Acta Cryst.* D**55**, 484–491.

Kleywegt, G. J. (1996). *Use of non-crystallographic symmetry in protein structure refinement*. *Acta Cryst.* D**52**, 842–857.

Kraut, J. (1961). *The crystal structure of 2-amino-ethanol phosphate. Acta Cryst.* **14**, 1146–1152.

Lamzin, V. S., Perrakis, A. & Wilson, K. S. (2001). *The ARP/wARP suite for automated construction and refinement of protein models*. In *International Tables for Crystallography*, Vol. F, *Crystallography of Biological Macromolecules*, edited by M. G. Rossmann & E. Arnold, Section 25.2.5. Dordrecht: Kluwer Academic Publishers.

Lattman, E. E. (1972). *Optimal sampling of the rotation function. Acta Cryst.* B**28**, 1065–1068.

Lattman, E. E. & Love, W. E. (1970). *A rotational search procedure for detecting a known molecule in a crystal. Acta Cryst.* B**26**, 1854–1857.

Lentz, P. J. Jr, Strandberg, B., Unge, T., Vaara, I., Borell, A., Fridborg, K. & Petef, G. (1976). *The determination of the heavy-atom substitution sites in the satellite tobacco necrosis virus. Acta Cryst.* B**32**, 2979–2983.

Leslie, A. G. W. (1987). *A reciprocal-space method for calculating a molecular envelope using the algorithm of B.C. Wang*. *Acta Cryst.* A**43**, 134–136.

Lifchitz, A. (1983). *On the choice of the model cell and the integration volume in the use of the rotation function. Acta Cryst.* A**39**, 130–139.

Liljas, L., Unge, T., Jones, T. A., Fridborg, K., Lövgren, S., Skoglund, U. & Strandberg, B. (1982). *Structure of satellite tobacco necrosis virus at 3.0 Å resolution. J. Mol. Biol.* **159**, 93–108.

Lipson, H. & Cochran, W. (1966). *The Determination of Crystal Structures.* Ithaca: Cornell University Press.

Litvin, D. B. (1975). *The molecular replacement method. I. The rotation function problem, application to bovine liver catalase and STNV*. *Acta Cryst.* A**31**, 407–416.

Lu, G. (1999). *FINDNCS: a program to detect non-crystallographic symmetries in protein crystals from heavy-atom sites*. *J. Appl. Cryst.* **32**, 365–368.

Lunin, V. Y. (1993). *Electron-density histograms and the phase problem*. *Acta Cryst.* D**49**, 90–99.

Lunin, V. Y., Lunina, N. L., Petrova, T. E., Skovoroda, T. P., Urzhumtsev, A. G. & Podjarny, A. D. (2000). *Low-resolution ab initio phasing: problems and advances*. *Acta Cryst.* D**56**, 1223–1232.

Luo, M., Vriend, G., Kamer, G. & Rossmann, M. G. (1989). *Structure determination of Mengo virus*. *Acta Cryst.* B**45**, 85–92.

Luzzati, V. (1953). *Résolution d'une structure cristalline lorsque les positions d'une partie des atomes sont connues: traitement statistique. Acta Cryst.* **6**, 142–152.

McKenna, R., Xia, D., Willingmann, P., Ilag, L. L. & Rossmann, M. G. (1992). *Structure determination of the bacteriophage φX174*. *Acta Cryst.* B**48**, 499–511.

McLachlan, D. Jr & Harker, D. (1951). *Finding the signs of the F's from the shifted Patterson product. Proc. Natl Acad. Sci. USA*, **37**, 846–849.

Main, P. (1967). *Phase determination using non-crystallographic symmetry. Acta Cryst.* **23**, 50–54.

Main, P. & Rossmann, M. G. (1966). *Relationships among structure factors due to identical molecules in different crystallographic environments. Acta Cryst.* **21**, 67–72.

Matthews, B. W. (1966). *The determination of the position of anomalously scattering heavy atom groups in protein crystals. Acta Cryst.* **20**, 230–239.

Matthews, B. W. & Czerwinski, E. W. (1975). *Local scaling: a method to reduce systematic errors in isomorphous replacement and anomalous scattering measurements. Acta Cryst.* A**31**, 480–487.

Matthews, B. W., Sigler, P. B., Henderson, R. & Blow, D. M. (1967). *Three-dimensional structure of tosyl-α-chymotrypsin. Nature (London),* **214**, 652–656.

Menzer, G. (1949). *Über die mehrdeutigkeit der Kristallstrukturbestimmung. Z. Naturforsch. Teil A*, **4**, 11–21.

Mighell, A. D. & Jacobson, R. A. (1963). *Analysis of three-dimensional Patterson maps using vector verification. Acta Cryst.* **16**, 443–445.

Miller, S. T., Hogle, J. M. & Filman, D. J. (2001). *Ab initio phasing of high-symmetry macromolecular complexes: successful phasing of authentic poliovirus data to 3.0 Å resolution. J. Mol. Biol.* **307**, 499–512.

Moncrief, J. W. & Lipscomb, W. N. (1966). *Structure of leurocristine methiodide dihydrate by anomalous scattering methods; relation to leurocristine (vincristine) and vincaleukoblastine (vinblastine). Acta Cryst.* **21**, 322–331.

Moras, D., Comarmond, M. B., Fischer, J., Weiss, R., Thierry, J. C., Ebel, J. P. & Giegé, R. (1980). *Crystal structure of yeast tRNA^{Asp}. Nature (London)*, **288**, 669–674.

Muirhead, H., Cox, J. M., Mazzarella, L. & Perutz, M. F. (1967). *Structure and function of haemoglobin. III. A three-dimensional Fourier synthesis of human deoxyhaemoglobin at 5.5 Å resolution. J. Mol. Biol.* **28**, 117–156.

Murthy, M. R. N., Reid, T. J. III, Sicignano, A., Tanaka, N. & Rossmann, M. G. (1981). *Structure of beef liver catalase. J. Mol. Biol.* **152**, 465–499.

Nagem, R. A. P., Polikarpov, I. & Dauter, Z. (2003). *Phasing on rapidly soaked ions. Methods Enzymol.* **374**, 120–137.

Navaza, J. (1987). *On the fast rotation function. Acta Cryst.* A**43**, 645–653.

Navaza, J. (1993). *On the computation of the fast rotation function. Acta Cryst.* D**49**, 588–591.

Navaza, J. (1994). *AMoRe: an automated package for molecular replacement. Acta Cryst.* A**50**, 157–163.

Navaza, J. (2001*a*). *Implementation of molecular replacement in AMoRe. Acta Cryst.* D**57**, 1367–1372.

Navaza, J. (2001*b*). *Rotation functions*. In *International Tables for Crystallography*, Vol. F, *Crystallography of Biological Macromolecules*, edited by M. G. Rossmann & E. Arnold, ch. 13.2. Dordrecht: Kluwer Academic Publishers.

Navaza, J., Panepucci, E. H. & Martin, C. (1998). *On the use of strong Patterson function signals in many-body molecular replacement. Acta Cryst.* D**54**, 817–821.

Navaza, J. & Vernoslova, E. (1995). *On the fast translation functions for molecular replacement. Acta Cryst.* A**51**, 445–449.

Nieh, Y.-P. & Zhang, K. Y. J. (1999). *A two-dimensional histogram-matching method for protein phase refinement and extension. Acta Cryst.* D**55**, 1893–1900.

Nixon, P. E. (1978). *Overlapping Patterson peaks and direct methods: the structure of prostratin. Acta Cryst.* A**34**, 450–453.

Nordman, C. E. (1966). *Vector space search and refinement procedures. Trans. Am. Crystallogr. Assoc.* **2**, 29–38.

Nordman, C. E. (1972). *An application of vector space search methods to the Patterson function of myoglobin. Acta Cryst.* A**28**, 134–143.

Nordman, C. E. (1980*a*). *Vector-space Patterson search and other stored-function sampling procedures.* In *Computing in Crystallography*, edited by R. Diamond, S. Ramaseshan & K. Venkatesan, pp. 5.01–5.13. Bangalore: Indian Academy of Sciences.

Nordman, C. E. (1980*b*). *Procedures for detection and idealization of non-crystallographic symmetry with application to phase refinement of the satellite tobacco necrosis virus structure. Acta Cryst.* A**36**, 747–754.

Nordman, C. E. & Nakatsu, K. (1963). *Interpretation of the Patterson function of crystals containing a known molecular fragment. The structure of an Alstonia alkaloid. J. Am. Chem. Soc.* **85**, 353–354.

Nordman, C. E. & Schilling, J. W. (1970). *Calculation and use of vector overlap weights in Patterson search and refinement.* In *Crystallographic Computing*, edited by F. R. Ahmed, S. R. Hall & C. P. Huber, pp. 110–114. Copenhagen: Munksgaard.

North, A. C. T. (1965). *The combination of isomorphous replacement and anomalous scattering data in phase determination of non-centrosymmetric reflexions. Acta Cryst.* **18**, 212–216.

Okaya, Y., Saito, Y. & Pepinsky, R. (1955). *New method in X-ray crystal structure determination involving the use of anomalous dispersion. Phys. Rev.* **98**, 1857–1858.

Pannu, N. S., Murshudov, G. N., Dodson, E. J. & Read, R. J. (1998). *Incorporation of prior phase information strengthens maximum-likelihood structure refinement. Acta Cryst.* D**54**, 1285–1294.

Patterson, A. L. (1934*a*). *A Fourier series representation of the average distribution of scattering power in crystals. Phys. Rev.* **45**, 763.

Patterson, A. L. (1934*b*). *A Fourier series method for the determination of the components of interatomic distances in crystals. Phys. Rev.* **46**, 372–376.

Patterson, A. L. (1935). *A direct method for the determination of the components of interatomic distances in crystals. Z. Kristallogr.* **90**, 517–542.

Patterson, A. L. (1939). *Homometric structures. Nature (London)*, **143**, 939–940.

Patterson, A. L. (1944). *Ambiguities in the X-ray analysis of crystal structures. Phys. Rev.* **65**, 195–201.

Patterson, A. L. (1949). *An alternative interpretation for vector maps. Acta Cryst.* **2**, 339–340.

Pauling, L. & Shappell, M. D. (1930). *The crystal structure of bixbyite and the C-modification of the sesquioxides. Z. Kristallogr.* **75**, 128–142.

Pepinsky, R., Okaya, Y. & Takeuchi, Y. (1957). *Theory and application of the $P_s(u)$ function and anomalous dispersion in direct determination of structures and absolute configuration in non-centric crystals. Acta Cryst.* **10**, 756.

Perrakis, A., Sixma, T. K., Wilson, K. S. & Lamzin, V. S. (1997). *wARP: improvement and extension of crystallographic phases by weighted averaging of multiple-refined dummy atomic models. Acta Cryst.* D**53**, 448–455.

Perutz, M. F. (1954). *The structure of haemoglobin. III. Direct determination of the molecular transform. Proc. R. Soc. London Ser. A*, **225**, 264–286.

Perutz, M. F. (1956). *Isomorphous replacement and phase determination in non-centrosymmetric space groups. Acta Cryst.* **9**, 867–873.

Phillips, D. C. (1966). *Advances in protein crystallography.* In *Advances in Structure Research by Diffraction Methods*, Vol. 2, edited by R. Brill & R. Mason, pp. 75–140. New York: John Wiley.

Poljak, R. J. (1963). *Heavy-atom attachment to crystalline lysozyme. J. Mol. Biol.* **6**, 244–246.

Rabinovich, D. & Shakked, Z. (1984). *A new approach to structure determination of large molecules by multi-dimensional search methods. Acta Cryst.* A**40**, 195–200.

Rae, A. D. (1977). *The use of structure factors to find the origin of an oriented molecular fragment. Acta Cryst.* A**33**, 423–425.

Ramachandran, G. N. & Raman, S. (1959). *Syntheses for the deconvolution of the Patterson function. Part I. General principles. Acta Cryst.* **12**, 957–964.

Ramagopal, U. A., Dauter, M. & Dauter, Z. (2003). *Phasing on anomalous signal of sulfurs: what is the limit? Acta Cryst.* D**59**, 1020–1027.

Raman, S. (1966). *Patterson functions and vector sets. Trans. Am. Crystallogr. Assoc.* **2**, 10–16.

Raman, S. & Lipscomb, W. N. (1961). *Two classes of functions for the location of heavy atoms and for solution of crystal structures. Z. Kristallogr.* **116**, 314–327.

Ramaseshan, S. & Abrahams, S. C. (1975). Editors. *Anomalous Scattering.* Copenhagen: Munksgaard.

Rao, S. N., Jih, J. H. & Hartsuck, J. A. (1980). *Rotation-function space groups. Acta Cryst.* A**36**, 878–884.

Rayment, I. (1983). *Molecular replacement method at low resolution: optimum strategy and intrinsic limitations as determined by calculations on icosahedral virus models. Acta Cryst.* A**39**, 102–116.

Rayment, I., Baker, T. S. & Caspar, D. L. D. (1983). *A description of the techniques and application of molecular replacement used to determine the structure of polyoma virus capsid at 22.5 Å resolution. Acta Cryst.* B**39**, 505–516.

Rayment, I., Baker, T. S., Caspar, D. L. D. & Murakami, W. T. (1982). *Polyoma virus capsid structure at 22.5 Å resolution. Nature (London)*, **295**, 110–115.

Rayment, I., Johnson, J. E., Suck, D., Akimoto, T. & Rossmann, M. G. (1978). *An 11 Å resolution electron density map of southern bean mosaic virus. Acta Cryst.* B**34**, 567–578.

Read, R. J. (2001*a*). *Model phases: probabilities, bias and maps*. In *International Tables for Crystallography*, Vol. F, *Crystallography of*

Biological Macromolecules, edited by M. G. Rossmann & E. Arnold, ch. 15.2. Dordrecht: Kluwer Academic Publishers.

Read, R. J. (2001*b*). *Pushing the boundaries of molecular replacement with maximum likelihood*. *Acta Cryst.* D**57**, 1373–1382.

Rees, D. C. & Lewis, M. (1983). *Incorporation of experimental phases in a restrained least-squares refinement*. *Acta Cryst.* A**39**, 94–97.

Refaat, L. S., Tate, C. & Woolfson, M. M. (1996). *Direct-space methods in phase extension and phase refinement. IV. The double-histogram method*. *Acta Cryst.* D**52**, 252–256.

Ren, J., Esnouf, R., Garman, E., Somers, D., Ross, C., Kirby, I., Keeling, J., Darby, G., Jones, Y., Stuart, D. & Stammers, D. (1995). *High resolution structures of HIV-1 RT from four RT-inhibitor complexes*. *Nature Struct. Biol.* **2**, 293–302.

Robertson, J. M. (1935). *An X-ray study of the structure of phthalocyanines. Part I. The metal-free, nickel, copper, and platinum compounds. J. Chem. Soc.* pp. 615–621.

Robertson, J. M. (1936). *An X-ray study of the phthalocyanines. Part II. Quantitative structure determination of the metal-free compound. J. Chem. Soc.* pp. 1195–1209.

Robertson, J. M. (1951). *Interpretation of the Patterson synthesis: rubidium benzyl penicillin. Acta Cryst.* **4**, 63–66.

Robertson, J. M. & Woodward, I. (1937). *An X-ray study of the phthalocyanines. Part III. Quantitative structure determination of nickel phthalocyanine. J. Chem. Soc.* pp. 219–230.

Rogers, D. (1951). *New methods of direct structure determination using modified Patterson maps. Research*, **4**, 295–296.

Rossmann, M. G. (1960). *The accurate determination of the position and shape of heavy-atom replacement groups in proteins. Acta Cryst.* **13**, 221–226.

Rossmann, M. G. (1961*a*). *The position of anomalous scatterers in protein crystals. Acta Cryst.* **14**, 383–388.

Rossmann, M. G. (1961*b*). *Application of the Buerger minimum function to protein structures.* In *Computing Methods and the Phase Problem in X-ray Crystal Analysis*, edited by R. Pepinsky, J. M. Robertson & J. C. Speakman, pp. 252–265. Oxford: Pergamon Press.

Rossmann, M. G. (1972). *The Molecular Replacement Method.* New York: Gordon & Breach.

Rossmann, M. G. (1990). *The molecular replacement method. Acta Cryst.* A**46**, 73–82.

Rossmann, M. G. (2001). *Molecular replacement – historical background*. *Acta Cryst.* D**57**, 1360–1366.

Rossmann, M. G. & Arnold, E. (2001*a*). Editors. *International Tables for Crystallography*, Vol. F, *Crystallography of Biological Macromolecules*. Dordrecht: Kluwer Academic Publishers.

Rossmann, M. G. & Arnold, E. (2001*b*). *Noncrystallographic symmetry averaging of electron density for molecular-replacement phase refinement and extension*. In *International Tables for Crystallography*, Vol. F, *Crystallography of Biological Macromolecules*, edited by M. G. Rossmann & E. Arnold, ch. 13.4. Dordrecht: Kluwer Academic Publishers.

Rossmann, M. G., Arnold, E., Erickson, J. W., Frankenberger, E. A., Griffith, J. P., Hecht, H. J., Johnson, J. E., Kamer, G., Luo, M., Mosser, A. G., Rueckert, R. R., Sherry, B. & Vriend, G. (1985). *Structure of a human common cold virus and functional relationship to other picornaviruses. Nature (London)*, **317**, 145–153.

Rossmann, M. G. & Blow, D. M. (1961). *The refinement of structures partially determined by the isomorphous replacement method. Acta Cryst.* **14**, 641–647.

Rossmann, M. G. & Blow, D. M. (1962). *The detection of sub-units within the crystallographic asymmetric unit. Acta Cryst.* **15**, 24–31.

Rossmann, M. G. & Blow, D. M. (1963). *Determination of phases by the conditions of non-crystallographic symmetry. Acta Cryst.* **16**, 39–45.

Rossmann, M. G., Blow, D. M., Harding, M. M. & Coller, E. (1964). *The relative positions of independent molecules within the same asymmetric unit. Acta Cryst.* **17**, 338–342.

Rossmann, M. G., Ford, G. C., Watson, H. C. & Banaszak, L. J. (1972). *Molecular symmetry of glyceraldehyde-3-phosphate dehydrogenase. J. Mol. Biol.* **64**, 237–249.

Rossmann, M. G. & Henderson, R. (1982). *Phasing electron diffraction amplitudes with the molecular replacement method. Acta Cryst.* A**38**, 13–20.

Rossmann, M. G., McKenna, R., Tong, L., Xia, D., Dai, J.-B., Wu, H., Choi, H.-K. & Lynch, R. E. (1992). *Molecular replacement real-space averaging*. *J. Appl. Cryst.* **25**, 166–180.

Roversi, P., Blanc, E., Vonrhein, C., Evans, G. & Bricogne, G. (2000). *Modelling prior distributions of atoms for macromolecular refinement and completion*. *Acta Cryst.* D**56**, 1316–1323.

Sasada, Y. (1964). *The differential rotation function. Acta Cryst.* **17**, 611–612.

Sayre, D. (1952). *The squaring method: a new method for phase determination*. *Acta Cryst.* **5**, 60–65.

Schevitz, R. W., Podjarny, A. D., Zwick, M., Hughes, J. J. & Sigler, P. B. (1981). *Improving and extending the phases of medium- and low-resolution macromolecular structure factors by density modification. Acta Cryst.* A**37**, 669–677.

Schiltz, M., Fourme, R. & Prange, T. (2003). *Use of noble gases xenon and krypton as heavy atoms in protein structure determination*. *Methods Enzymol.* **374**, 83–119.

Schuller, D. J. (1996). *MAGICSQUASH: more versatile non-crystallographic averaging with multiple constraints*. *Acta Cryst.* D**52**, 425–434.

Sheriff, S., Klei, H. E. & Davis, M. E. (1999). *Implementation of a six-dimensional search using the AMoRe translation function for difficult molecular-replacement problems*. *J. Appl. Cryst.* **32**, 98–101.

Shoemaker, D. P., Donohue, J., Schomaker, V. & Corey, R. B. (1950). *The crystal structure of L_8-threonine. J. Am. Chem. Soc.* **72**, 2328–2349.

Sim, G. A. (1961). *Aspects of the heavy-atom method.* In *Computing Methods and the Phase Problem in X-ray Crystal Analysis*, edited by R. Pepinsky, J. M. Robertson & J. C. Speakman, pp. 227–235. Oxford: Pergamon Press.

Simonov, V. I. (1965). *Calculation of the phases of the structure amplitudes by Fourier transformation of the sum, product and minimum functions. Proc. Indian Acad. Sci.* A**62**, 213–223.

Simpson, A. A., Leiman, P. G., Tao, Y., He, Y., Badasso, M. O., Jardine, P. J., Anderson, D. L. & Rossmann, M. G. (2001). *Structure determination of the head-tail connector of bacteriophage φ29*. *Acta Cryst.* D**57**, 1260–1269.

Simpson, P. G., Dobrott, R. D. & Lipscomb, W. N. (1965). *The symmetry minimum function: high order image seeking functions in X-ray crystallography*. *Acta Cryst.* **18**, 169–179.

Singh, A. K. & Ramaseshan, S. (1966). *The determination of heavy atom positions in protein derivatives. Acta Cryst.* **21**, 279–280.

Smith, J. L., Hendrickson, W. A. & Addison, A. W. (1983). *Structure of trimeric haemerythrin. Nature (London)*, **303**, 86–88.

Speakman, J. C. (1949). *The crystal structures of the acid salts of some monobasic acids. Part I. Potassium hydrogen bisphenyl acetate. J. Chem. Soc.* pp. 3357–3365.

Stauffacher, C. V., Usha, R., Harrington, M., Schmidt, T., Hosur, M. V. & Johnson, J. E. (1987). *The structure of cowpea mosaic virus at 3.5 Å resolution.* In *Crystallography in Molecular Biology*, edited by D. Moras, J. Drenth, B. Strandberg, D. Suck & K. Wilson, pp. 293–308. New York, London: Plenum.

Steinrauf, L. K. (1963). *Two Fourier functions for use in protein crystallography*. *Acta Cryst.* **16**, 317–319.

Storoni, L. C., McCoy, A. J. & Read, R. J. (2004). *Likelihood-enhanced fast rotation functions*. *Acta Cryst.* D**60**, 432–438.

Stout, G. H. & Jensen, L. H. (1968). *X-ray Structure Determination.* New York: Macmillan.

Strahs, G. & Kraut, J. (1968). *Low-resolution electron-density and anomalous-scattering-density maps of Chromatium high-potential iron protein. J. Mol. Biol.* **35**, 503–512.

Tanaka, N. (1977). *Representation of the fast-rotation function in a polar coordinate system. Acta Cryst.* A**33**, 191–193.

Taylor, W. J. (1953). *Fourier representation of Buerger's image-seeking minimum function. J. Appl. Phys.* **24**, 662–663.

Terwilliger, T. C. (1999). *Reciprocal-space solvent flattening*. *Acta Cryst.* D**55**, 1863–1871.

Terwilliger, T. C. (2002*a*). *Rapid automatic NCS identification using heavy-atom substructures*. *Acta Cryst.* D**58**, 2213–2215.

Terwilliger, T. C. (2002*b*). *Statistical density modification with non-crystallographic symmetry*. *Acta Cryst.* D**58**, 2082–2086.

Terwilliger, T. C. (2003*a*). *Improving macromolecular atomic models at moderate resolution by automated iterative model building, statistical density modification and refinement*. *Acta Cryst.* D**59**, 1174–1182.

Terwilliger, T. C. (2003*b*). *SOLVE and RESOLVE: Automated structure solution and density modification*. *Methods Enzymol.* **374**, 22–37.

Terwilliger, T. C. (2003*c*). *Statistical density modification using local pattern matching*. *Acta Cryst.* D**59**, 1688–1701.

Terwilliger, T. C. (2004). *Using prime-and-switch phasing to reduce model bias in molecular replacement*. *Acta Cryst.* D**60**, 2144–2149.
Terwilliger, T. C. & Berendzen, J. (1999). *Automated MAD and MIR structure solution*. *Acta Cryst.* D**55**, 849–861.
Terwilliger, T. C. & Berendzen, J. (2001). *Automated MAD and MIR structure solution*. In *International Tables for Crystallography*, Vol. F, *Crystallography of Biological Macromolecules*, edited by M. G. Rossmann & E. Arnold, Section 14.2.2. Dordrecht: Kluwer Academic Publishers.
Terwilliger, T. C. & Eisenberg, D. (1983). *Unbiased three-dimensional refinement of heavy-atom parameters by correlation of origin-removed Patterson functions. Acta Cryst.* A**39**, 813–817.
Terwilliger, T. C., Kim, S.-H. & Eisenberg, D. (1987). *Generalized method of determining heavy-atom positions using the difference Patterson function*. *Acta Cryst.* A**43**, 1–5.
Tollin, P. (1966). *On the determination of molecular location. Acta Cryst.* **21**, 613–614.
Tollin, P. (1969). *A comparison of the Q-functions and the translation function of Crowther and Blow*. *Acta Cryst.* A**25**, 376–377.
Tollin, P. & Cochran, W. (1964). *Patterson function interpretation for molecules containing planar groups. Acta Cryst.* **17**, 1322–1324.
Tollin, P., Main, P. & Rossmann, M. G. (1966). *The symmetry of the rotation function. Acta Cryst.* **20**, 404–407.
Tollin, P. & Rossmann, M. G. (1966). *A description of various rotation function programs. Acta Cryst.* **21**, 872–876.
Tong, L. (1993). *REPLACE, a suite of computer programs for molecular-replacement calculations*. *J. Appl. Cryst.* **26**, 748–751.
Tong, L. (1996*a*). *Combined molecular replacement*. *Acta Cryst.* A**52**, 782–784.
Tong, L. (1996*b*). *The locked translation function and other applications of a Patterson correlation function*. *Acta Cryst.* A**52**, 476–479.
Tong, L. (2001*a*). *How to take advantage of non-crystallographic symmetry in molecular replacement: 'locked' rotation and translation functions*. *Acta Cryst.* D**57**, 1383–1389.
Tong, L. (2001*b*). *Translation functions*. In *International Tables for Crystallography*, Vol. F, *Crystallography of Biological Macromolecules*, edited by M. G. Rossmann & E. Arnold, ch. 13.3. Dordrecht: Kluwer Academic Publishers.
Tong, L., Qian, C., Davidson, W., Massariol, M.-J., Bonneau, P. R., Cordingley, M. G. & Lagacé, L. (1997). *Experiences from the structure determination of human cytomegalovirus protease*. *Acta Cryst.* D**53**, 682–690.
Tong, L. & Rossmann, M. G. (1990). *The locked rotation function. Acta Cryst.* A**46**, 783–792.
Tong, L. & Rossmann, M. G. (1993). *Patterson-map interpretation with noncrystallographic symmetry*. *J. Appl. Cryst.* **26**, 15–21.
Tong, L. & Rossmann, M. G. (1995). *Reciprocal-space molecular-replacement averaging*. *Acta Cryst.* D**51**, 347–353.
Tong, L. & Rossmann, M. G. (1997). *Rotation function calculations with GLRF program*. *Methods Enzymol.* **276**, 594–611.
Tronrud, D. E. & Ten Eyck, L. F. (2001). *The TNT refinement package*. In *International Tables for Crystallography*, Vol. F, *Crystallography of Biological Macromolecules*, edited by M. G. Rossmann & E. Arnold, Section 25.2.4. Dordrecht: Kluwer Academic Publishers.
Tsao, J., Chapman, M. S. & Rossmann, M. G. (1992). *Ab initio phase determination for viruses with high symmetry: a feasibility study*. *Acta Cryst*. A**48**, 293–301.
Urzhumtsev, A. & Podjarny, A. (1995). *On the solution of the molecular-replacement problem at very low resolution: Application to large complexes*. *Acta Cryst.* D**51**, 888–895.
Vagin, A. & Teplyakov, A. (2000). *An approach to multi-copy search in molecular replacement*. *Acta Cryst.* D**56**, 1622–1624.
Wang, B. C. (1985). *Resolution of phase ambiguity in macromolecular crystallography*. *Methods Enzymol.* **115**, 90–112.
Weeks, C. M., Adams, P. D., Berendzen, J., Brunger, A. T., Dodson, E. J., Grosse-Kunstleve, R. W., Schneider, T. R., Sheldrick, G. M., Terwilliger, T. C., Turkenburg, M. G. & Uson, I. (2003). *Automatic solution of heavy-atom substructures*. *Methods Enzymol.* **374**, 37–83.
Wilson, A. J. C. (1942). *Determination of absolute from relative X-ray intensity data. Nature (London)*, **150**, 151–152.
Wilson, I. A., Skehel, J. J. & Wiley, D. C. (1981). *Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 Å resolution. Nature (London)*, **289**, 366–373.
Woolfson, M. M. (1956). *An improvement of the 'heavy-atom' method of solving crystal structures. Acta Cryst.* **9**, 804–810.
Woolfson, M. M. (1970). *An Introduction to X-ray Crystallography*. London: Cambridge University Press.
Wrinch, D. M. (1939). *The geometry of discrete vector maps. Philos. Mag.* **27**, 98–122.
Wunderlich, J. A. (1965). *A new expression for sharpening Patterson functions. Acta Cryst.* **19**, 200–202.
Yang, C., Pflugrath, J. W., Courville, D. A., Stence, C. N. & Ferrara, J. D. (2003). *Away from the edge: SAD phasing from the sulfur anomalous signal measured in-house with chromium radiation*. *Acta Cryst.* D**59**, 1943–1957.
Yeates, T. O. & Rini, J. M. (1990). *Intensity-based domain refinement of oriented but unpositioned molecular replacement models*. *Acta Cryst*. A**46**, 352–359.
Zhang, K. Y. J. (1993). *SQUASH – combining constraints for macromolecular phase refinement and extension*. *Acta Cryst.* D**49**, 213–222.
Zhang, K. Y. J., Cowtan, K. D. & Main, P. (2001). *Phase improvement by iterative density modification*. In *International Tables for Crystallography*, Vol. F, *Crystallography of Biological Macromolecules*, edited by M. G. Rossmann & E. Arnold, ch 15.1. Dordrecht: Kluwer Academic Publishers.