

2. RECIPROCAL SPACE IN CRYSTAL-STRUCTURE DETERMINATION

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		$n = 3$

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 ($^\circ$)		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

(Stauffer *et al.*, 1987)] and enzyme [catalase (Murthy *et al.*, 1981)] heavy-atom difference Pattersons. 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

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

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

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

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 interparticle 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.

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)

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

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

2.3. PATTERSON AND MOLECULAR-REPLACEMENT TECHNIQUES

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

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

The elements of $[\mathbf{C}]$ will depend on three rotation angles $(\theta_1, \theta_2, \theta_3)$. Thus, R is a function of these three angles. Alternatively, the matrix $[\mathbf{C}]$ 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

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

and, hence, by substitution in (2.3.6.1)

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

where

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

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

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

where $\theta = 2\pi H R$ and $H = \mathbf{h} + \mathbf{p}[\mathbf{C}]$. 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 $\mathbf{h}' = [\mathbf{C}^T]\mathbf{p}$ 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 $[\mathbf{C}]$. 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

2. RECIPROCAL SPACE IN CRYSTAL-STRUCTURE DETERMINATION

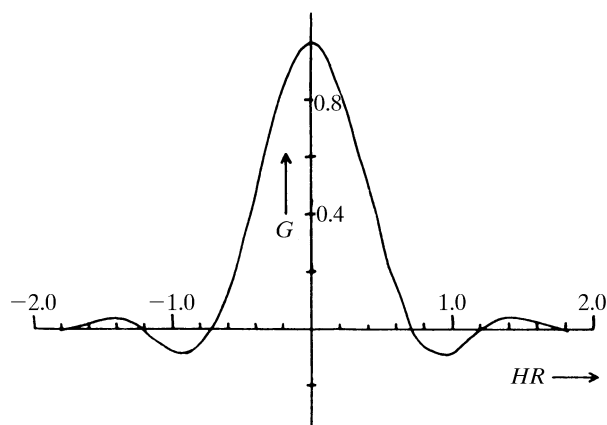


Fig. 2.3.6.1. Shape of the interference function G for a spherical envelope of radius R at a distance H from the reciprocal-space origin. [Reprinted from Rossmann & Blow (1962).]

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

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

or

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

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 $\sum_{\mathbf{p}} |\mathbf{F}_{\mathbf{p}}|^2 G_{\mathbf{h}\mathbf{p}}$ 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).

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 \mathbf{x} , these can be orthogonalized by a matrix $[\beta]$ to give the coordinates \mathbf{X} in units of length (Fig. 2.3.6.2); that is,

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

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

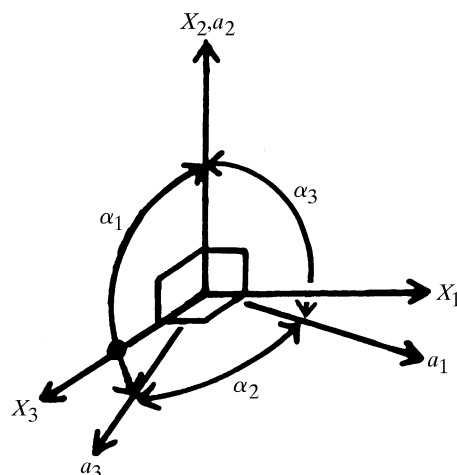


Fig. 2.3.6.2. Relationships of the orthogonal axes X_1, X_2, X_3 to the crystallographic axes a_1, a_2, a_3 . [Reprinted from Rossmann & Blow (1962).]

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

where ρ 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

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

Thus, by substitution,

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

and by comparison with (2.3.6.2) it follows that

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

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

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

and

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

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

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 $[\rho]$ in terms of Eulerian angles $\theta_1, \theta_2, \theta_3$:

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

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