

## 2.3. PATTERSON AND MOLECULAR-REPLACEMENT TECHNIQUES

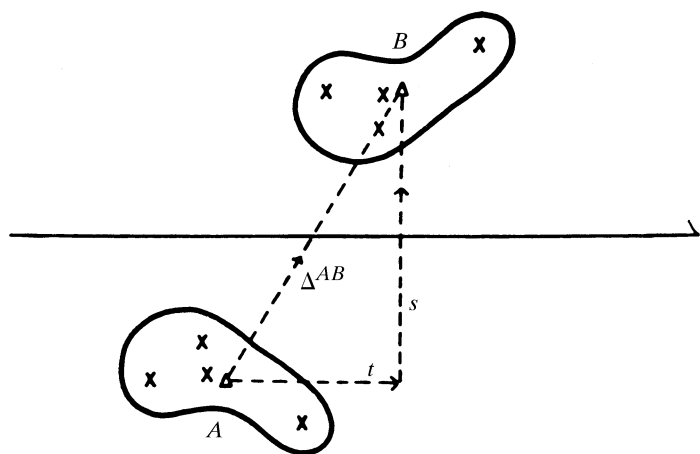


Fig. 2.3.7.1. Crosses represent atoms in a two-dimensional model structure. The triangles are the points chosen as approximate centres of molecules *A* and *B*.  $\Delta^{AB}$  has components  $t$  and  $s$  parallel and perpendicular, respectively, to the screw rotation axis. [Reprinted from Rossmann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

$$T(\mathbf{S}_x, \mathbf{S}_x') = \frac{2}{V_h V_p} \sum_{\mathbf{h}} \sum_{\mathbf{p}} |\mathbf{F}_h| |\mathbf{F}_p| G_{\mathbf{h}\mathbf{p}} \times \cos[\alpha_{\mathbf{h}} + \alpha_{\mathbf{p}} - 2\pi(\mathbf{h} \cdot \mathbf{S}_x + \mathbf{p} \cdot \mathbf{S}_x')]. \quad (2.3.7.1)$$

## 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

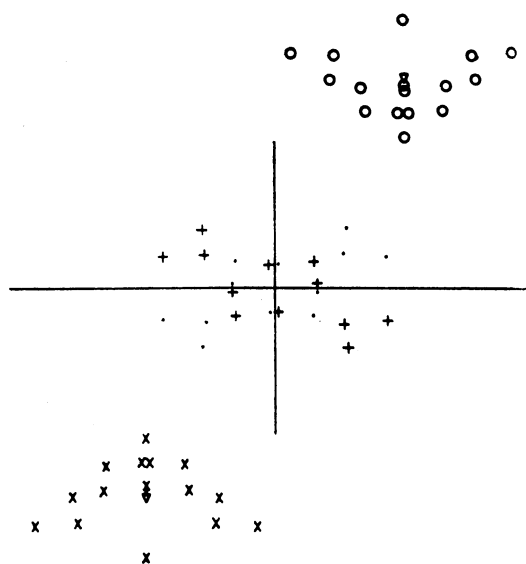


Fig. 2.3.7.2. Vectors arising from the structure in Fig. 2.3.7.1. The self-vectors of molecules *A* and *B* are represented by + and ·; the cross-vectors from molecules *A* to *B* and *B* to *A* by × and ○. Triangles mark the position of  $+\Delta^{AB}$  and  $-\Delta^{AB}$ . [Reprinted from Rossmann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

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),

$$T(\mathbf{S}) = \frac{2}{\sqrt{2}} \sum_{\mathbf{h}} \sum_{\mathbf{p}} |\mathbf{F}_h|^2 |\mathbf{F}_p|^2 G_{\mathbf{h}\mathbf{p}} \cos[2\pi(\mathbf{h} - \mathbf{p}) \cdot \mathbf{S}]. \quad (2.3.7.2)$$

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 ( $\mathbf{p}$ ) of an unknown structure and placed at  $\mathbf{S}$  with respect to a defined origin, then a suitable translation function is

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

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,

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

## 2. RECIPROCAL SPACE IN CRYSTAL-STRUCTURE DETERMINATION

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

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

which are the coefficients of the molecular transform for the known molecule placed into the  $n$ th asymmetric unit of the  $\mathbf{p}$  cell. Thus

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

or

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

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

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

and then from (2.3.7.3)

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

which is a Fourier summation with known coefficients  $\{|\mathbf{F}_{\mathbf{p}, \text{obs}}|^2 A_{p,n} A_{p,m} \times \exp[i(\gamma_n - \gamma_m)]\}$  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  $\mathbf{S}$  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  $\mathbf{p}$  cell is monoclinic with its unique axis parallel to  $\mathbf{b}$  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  $\mathbf{F}_M$  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)

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

where  $[\mathbf{C}]$  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  $\mathbf{h}$  cell, into which the known molecule was placed, has the same dimensions as the  $\mathbf{p}$  cell.

$R$ -factor calculations are sometimes used to determine the position of a known molecular fragment in an unknown cell, particularly if only one parameter is being searched. Such calculations are computationally less convenient than the Fourier methods described above, but can be more sensitive. All these methods can be improved by simultaneous consideration of packing requirements of the molecular fragments (Harada *et al.*, 1981; Hendrickson & Ward, 1976; Rabinovich & Shakked, 1984). Indeed, packing considerations can frequently limit the search volume very considerably.

### 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  $\mathbf{S}_x = \mathbf{S}_y = \mathbf{S}$  and the function (2.3.7.1) now becomes

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

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

## 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 ( $\mathbf{h}$ ), within which coordinate positions shall be designated by  $\mathbf{x}$ . 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 ( $\mathbf{p}$ ) into which the same molecule has been placed  $N$  times related by the crystallographic symmetry operators  $[\mathbf{C}_1], \mathbf{d}_1; [\mathbf{C}_2], \mathbf{d}_2; \dots; [\mathbf{C}_N], \mathbf{d}_N$ . 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  $n$ th asymmetric unit of the  $\mathbf{p}$  crystal such that

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

where

$$\mathbf{y}_n = [\mathbf{C}_n] \mathbf{y}_1 + \mathbf{d}_n. \quad (2.3.8.2)$$

From the definition of a structure factor,

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

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

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

Now let the molecule in the  $\mathbf{h}$  crystal be related to the molecule in the first asymmetric unit of the  $\mathbf{p}$  crystal by the noncrystallographic symmetry operation

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